Soc Sci

Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE

THE PSYCHOMETRIC SOCIETY - ORGANIZED IN 1935

VOLUME 20
NUMBER 4
DECEMBER
1 9 5 5

PSYCHOMETRIKA, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

DECEMBER, 1955, VOLUME 20, NUMBER 4

Published by the Psychometric Society at 1407 Sherwood Avenue, Richmond 5, Virginia, Entered as second class matter at the Post Office of Richmond, Virginia. Editorial Office, Department of Psychology, The University of North Carolina, Chapel Hill, North Carolina.

Subscription Price: The regular subscription rate is $14.00 per volume. The subscriber receives each issue as it comes out, and, upon request, a second complete set for binding at the end of the year. All annual subscriptions start with the March issue and cover the calendar year. All back issues but two are available. Back issues are $14.00 per volume (one set only) or $3.50 per issue, with a 20 per cent discount to Psychometric Society Members. Members of the Psychometric Society pay annual dues of $7.00, of which $6.30 is in payment of a subscription to *Psychometrika.* Student members of the Psychometric Society pay annual dues of $4.00, of which $3.60 is in payment for the journal

Application for membership and student membership in the Psychometric Society, together with a check for dues for the calendar year in which application is made, should be sent to

LEONARD KOGAN
105 EAST 22ND STREET, NEW YORK 10, NEW YORK

Payments: All bills and orders are payable in advance.

Checks covering membership dues should be made payable to the *Psychometric Society.*

Checks covering regular subscriptions to *Psychometrika* (for non-members of the Psychometric Society) and back issue orders should be made payable to the *Psychometric Corporation.* All checks, notices of change of address, and business communications should be sent to

DAVID R. SAUNDERS, Treasurer, Psychometric Society and Psychometric Corporation
Educational Testing Service
P.O. Box 592
Princeton, New Jersey

Articles on the following subjects are published in *Psychometrika:*

(1) the development of quantitative rationale for the solution of psychological problems;
(2) general theoretical articles on quantitative methodology in the social and biological sciences;
(3) new mathematical and statistical techniques for the evaluation of psychological data;
(4) aids in the application of statistical techniques, such as nomographs, tables, work-sheet layouts, forms, and apparatus;
(5) critiques or reviews of significant studies involving the use of quantitative techniques.

The emphasis is to be placed on articles of type (1), in so far as articles of this type are available.

(*Continued on the back inside cover page*)

F.

Psychometrika

CONTENTS

VOLUME TWENTY **DECEMBER 1955** **NUMBER 4**

Louis Leon Thurstone

Louis Leon Thurstone

With the death of L. L. Thurstone on September 29, 1955, psychology lost one of its greatest, a unique figure on the psychological scene and one to whom psychologists will always be indebted. If any psychologist of the past quarter century deserved to be called Mr. Psychological Measurement, it was he. His major professional objective coincided with that of the Psychometric Society and of *Psychometrika*, both of which were founded under his leadership: The development of psychology as a quantitative, rational science. By virtue of his own contributions and his influence on others, psychology has taken long steps in the direction of fulfillment of this objective. No major aspect of the field of measurement was untouched by him.

Louis Leon Thurstone was born May 29, 1887, in Chicago, where in later years he spent the greater portion of his professional life and achieved his greatest distinction, at the University of Chicago. His parents were of native, Swedish stock, his father's occupations being, in turn, military instructor, Lutheran pastor, editor, and publisher. Owing to a mobile family life, Thurstone went to school in Illinois; Mississippi; Stockholm, Sweden; and Jamestown, New York. He attended Cornell University, where he specialized in engineering. Considering the few instances of which the writer has known in which psychologists have started from a base of engineering training, he has often thought that we should be better off if more psychologists had taken that educational route.

It was during his engineering-school days that the problem of the learning curve, and hence psychology, caught Thurstone's attention. On graduation, however, he was offered a position in the laboratory of Thomas A. Edison, where he spent the year of 1912. During the next two academic years, he taught engineering courses at the University of Minnesota, and there began his study of experimental psychology. Graduate work followed at the University of Chicago. In 1915 he accepted an assistantship in the new and active laboratory established by Walter V. Bingham at the Carnegie Institute of Technology. He received his doctorate from Chicago, with a dissertation on the learning curve. His academic rise at Carnegie was something of a record. Beginning with the rank of instructor in 1917, with a promotion each year he became professor and head of the department by 1920.

The year of 1923–24 was spent in Washington, D. C., with the Institute for Government Research, an agency devoted to the improvement of civil-service practices. From that time on, Thurstone had considerable influence, directly or indirectly, upon civil-service procedures.

After his marriage in the summer of 1924 to Thelma Gwinn, Thurstone assumed his professorship at the University of Chicago. In the course of time,

he had much to do with initiating and setting the pattern for the University's distinguished Board of Examinations. In 1938 he was honored with the appointment as Charles F. Grey Distinguished Service Professor. In 1948 he was Visiting Professor at the University of Frankfurt, and in the spring semester of 1953 at the University of Stockholm. He retired from Chicago in 1952, at which time he became Research Professor and Director of the Psychometric Laboratory at the University of North Carolina, which was his professional affiliation at the time of his death.

His unquestioned creative productivity can perhaps be attributed to certain traits that seem to stand out—his dissatisfaction with the status of psychology as he found it, his keen analytical ability, and his independence and originality of thought. These qualities showed themselves in a number of ways. For example, his originality was demonstrated relatively early. While a college undergraduate he developed a novel method for trisecting an angle and published a paper on it in the *Scientific American*. By the time he graduated he had developed a motion-picture camera and projector of unusual design. It was this that brought him to the attention of Edison. His independence of thinking showed itself in the fact that he did not read widely in the psychological literature, as he was quite willing to admit, and in his general choice of some of the less popular subjects to which to devote his energies.

His dissatisfactions were many. He was discontented with the state of affairs he found in connection with psychological testing, where a rapidly growing practice seemed to have little or no underlying theory. It was in 1924 that he published his only attempt at general psychological theory, in his book entitled *The Nature of Intelligence*. More clearly, the need for test theory led him into the development of quantitative formulations, culminating in his multiple-factor analysis. His impatience with what he considered to be the severe limitations of classical psychophysics led to his development of basic theory of psychological measurement as effected through human judgments. He regarded his law of comparative judgment as one of his most important achievements.

His performance as an analytical thinker was most clearly brought to the attention of the writer when, in the fall quarter of 1935, he was privileged to attend Thurstone's seminar. Whenever a new problem came up in the seminar, it was a revelation to observe him go to work on it. He very quickly went to the heart of the problem, singling out the important variables in a way that made the problem appear simple. Incidentally, attending the seminar then and later were a large number of students, some at the post-doctoral level, who have achieved distinction in the field of psychometrics in their own right.

Thurstone's many contributions are so well known to readers of *Psychometrika* that they need not be enumerated here. Although he has been held in great respect by those outside the field of measurement, many of his developments have not had the general impact that they should have had. Thurstone

did not found a school, nor did he deal in popular or spectacular subjects. He often spoke to "deaf ears" because so few psychologists were prepared by virtue of interest or mathematical preparation to listen. Perhaps his contributions that have gained the widest notice are his attitude-scaling methods and attitude scales, and his discovery of primary mental abilities and his tests of them. Of the many quantitative methods that he has provided, probably that of multiple-factor analysis stands first by a wide margin in potential usefulness and impact upon psychology in general.

It is not so well known that in the later years of his life his energies were very much devoted to performance tests of personality. This interest sprang in part from dissatisfactions with both projective tests and personality inventories. He was also working on material for a book that was to give a general treatment of psychological measurement and on revisions of his tests of primary mental abilities. It is hoped that many of these last efforts had gone far enough to reach publication.

Thurstone was always quite willing to acknowledge the assistance from his wife, Thelma. They collaborated on many studies and for years were jointly responsible for the American Council on Education college-aptitude examination. Their three sons are grown and started on their various professional careers. Besides the immediate family, Thurstone leaves behind a great many loyal and capable former students, who follow in the Thurstone-Chicago tradition, as well as many admirers around the world.

University of Southern California *J. P. Guilford*

PSYCHOMETRIKA—VOL. 20, NO. 4
DECEMBER, 1955

PSYCHOMETRIC THEORY: GENERAL AND SPECIFIC*

LEDYARD R TUCKER

PRINCETON UNIVERSITY
AND
EDUCATIONAL TESTING SERVICE

Fellow members of the Psychometric Society, and our colleagues, members of Division 5 of the American Psychological Association, I understand my mission tonight is to provide you with both humor and a mild message. Precedent has established an eminently high standard. Maybe I can explain the humbleness with which I approach this task by relating the account of an episode. This occurrence, like much illustrative data in *Psychometrika,* is fictitious, of course.

During my trip from the splendorous East Coast to the beauteous West Coast, I took vacation time to tour part of the magnificent country in between, as many of you, undoubtedly, did also. One afternoon I came upon a small tent village in a fairly remote mountain section. Not that such camps were so uncommon as to be remarkable in themselves; something about this particular camp caught my eye. In the center of this camp was a rustic conference table flanked on one side by a large blackboard mounted on cedar posts. This in itself was enough to raise a moderate interest, but my curiosity was intensely stimulated by another observation. Scribbled on this blackboard were a number of familiar symbols. Here was a sigma, there was a derivative, another place was a box labelled "factor matrix." How could this be? Needless to say I stopped to investigate. To my extreme pleasure I found myself among friends. A small group of Psychometricians had set up a summer seminar. There was nothing for me to do but to stay for dinner.

Before and during dinner we chatted about this and that. There were a number of requests for news such as: Were the Brooklyn Dodgers still running away with the race in the National League? Had Senator McCarthy started a new investigation? What were the leading plays in the summer theaters? After these matters had become settled, we sat back quietly for a period. This silence was ended by one member, whom, of course, I won't name, exclaiming: "Say, I've got one that should work. How about this? 751955." There was a short pause for ten seconds while the others seemed to stop and contemplate the proposition. Then there was loud and long laughter.

*Presidential Address to the Psychometric Society, September 5, 1955.

In the enthusiasm of the moment, another member called out "That would be like 683291." Again there was laughter. Later in these proceedings I was able to break in and ask for information. I was told that the group had found the off hours to become a bit boring after the first week of their conference. They had heard each other's jokes several times over and were tired of hiking and fishing. As an extra-curricular activity they had started to code the accumulation of humor. The numbers I had heard were such coded jokes. I decided to try this situation and so called out 10121492. There was not the least smile. Upon my inquiry of what was wrong, I wasn't given the hackneyed answer that "some could tell them and some could not." I had under-estimated our colleagues. They had not just gone about listing all jokes that they could think of and giving them a serial number in order of occurrence. They had gone about this activity like true Psychometricians. They had set up a system of analysis and classification of humor and had applied it to their sample of jokes. By use of systems of binary bits and addends of various other orders, they had established a system to emcompass all possible jokes. Each sub-sub-etc. class, consisting of the jokes in one cell of this multiple classification system, had a unique identifying code which incorporated as meaningful units the position on each mode of classification. These identifications were then transformed to numbers in the decimal system. By considering the implications of any of the almost infinite number of empty cells they might conclude that the situation was humorous and have generated a new joke. This was what I had seen going on. It was a life-saver for the group, for otherwise they would not have been able to put up with each other for more than the second week. My weakness was that I had not realized the nature of their system and of its subtleties. I had pulled several boners. No wonder my selection was not humorous. Worst of all I had picked the lowest classification on the built-in criterion. I had told them that this joke was not the least bit funny.

With the field of humor now having been thoroughly studied, we may turn our attentions to other areas. For a short period I will direct my attention to material related to the title of this address, Psychometric Theory: General and Specific. Here I am taking the privilege of expressing my opinions on a recommended course of action for Psychometrics. My major concern is to maximize the extent to which observations of psychological phenomena can be validly matched with expectations obtained from rational theories.

While not wanting to dwell upon the philosophy of science—not that I would be able to do so if I wished—there are several aspects of my statement that warrant short discussion. By observations I mean the records obtained by definite procedures during the course of phenomena or of the observation of phenomena. The observations are subject to operational definitions. By expectations I mean statements describing the expected nature of observa-

tions. Such statements might vary on a continuum between vagueness and definiteness. I wish to consider definite statements as constituting definite expectations. Any vagueness could be considered as the introduction of approximation in the expectation. Thus, we might have approximate expectations.

I might illustrate by reference to color matching. The theory of the color pyramid would indicate that a given combination of color disks in particular proportions on one color wheel would be matched by another group of color disks on another wheel in given proportions. When the colors of these disks and the various proportions are specified, a definite expectation is produced. A corresponding observation could be obtained in the laboratory. A subject would be shown the color wheel with the standard combination of disks. The second wheel could be supplied with the other set of disks and the subject be asked to adjust the proportions until the color produced matched the color of the first wheel. Our observation would contain the specifications of the situation and a record of the proportions established by the subject.

I have avoided saying that observations were to be represented by rational theories. To make this latter statement would permit use of theories with infinite degrees of freedom so that these theories could be adjusted to represent any collection of observations, past, present and future. These theories would have no power, however, in providing definite statements of expectations concerning observations other than those observations the theories were adjusted to represent. In contrast, I have chosen to emphasize the correspondence between theoretically derived expectations and observations. The power of a theory is directly related to the number of definite expectations it produces and inversely to the errors between these expectations and the observations.

Let us consider the role of general psychometric theories in maximizing the extent to which derived expectations correspond to observations. Two aspects of this maximization appear: First, there may be a maximum correspondence, that is, minimum errors between the expectations and the observations. Second, the theory may yield a broad range of expectations. This is the place where general theories hold promise. We will all, undoubtedly, agree that it is desirable to have a systematic theory which will provide valid expectations for a large number of kinds of observations for many phenomena. There is a necessity, though, that we know how to use this theory to arrive at these expectations. Note my boner at the summer seminar previously described. As a further illustration, consider an advertisement that appeared a while ago in one of the weekly Princeton papers:

> **Lost: A black female cocker spaniel. Will not answer to "Daisy," although that is her name.**

Unless each theory explicitly indicates phenomena and observations to which it is applicable, we may be in no better position than would someone searching for Daisy. I fear that in a number of instances we have mistaken vagueness for generality.

One may be interested in the relation between a method for analysis of data and a theory. In some senses a method of analysis might be considered as a general theory. A proper description of any such method includes a statement of the necessary conditions of situations in which the method might be employed. In a sense, however, we may contrast a method with a theory. This is in the specificity of definitions. If a construct might correspond to different kinds of observations in different situations, an ambiguity exists. A method that is applicable in several situations may be thought of as a theory in each situation. Each of these theories is separate from the other in that it includes the further definitions necessary to fit the method to the situation. Thus, a method may be considered as a family of theories. For example, multiple-factor analysis might be considered as a family of theories. Any time it is applied to a particular content area, it may be thought of as a theory. A method of analysis may be thought of as an abstraction from a number of situations with particular contents. In order to produce any particular theory from the family of theories represented by a method, it is necessary to specify particular contents.

We may be interested in classifying methods of analysis as more or less general. This classification may be taken to correspond to the number of different situations in which the method is applicable. A more general method would be applicable in a larger number of different situations. However, there may be a trend that the more general the method is, in this sense, the more extensive the definitions need be in order to make the method explicitly applicable to any one particular situation. Thus, the more general methods may be further from theories.

If a theory is to specify the observations and phenomena with which it deals, how then can a theory be general? We may wish to indicate the extensiveness of a theory's applicability by the contrasting terms: individual versus general. In case a theory concerns a large number of observables in a large variety of situations, the theory could be termed a relatively general theory. In case a theory concerns only a limited situation, the theory could be termed individual, or specific, to this situation. Note the particular usage of the word specific as synonymous to individual in this context. Both a general and an individual, or specific, theory should specify the observables with which it deals as well as situations or a situation in which the theory applies.

Although one might hope that a collection of individual theories could be amalgamated into a more general theory, there is some possibility that the individual theories may be too divergent to provide a basis for this

operation. Many individual theories arise from applications of psychometric methods to problems of applied psychology. The particular situations dealt with are defined by the needs of particular institutions. While this activity is a highly important aspect of psychometrics, its productivity of general theory is problematic. There is considerable danger that the practicing psychometrician may find himself in the kind of position illustrated by the following advertisement, which has also appeared in a weekly Princeton paper:

> Aspiring Artist: Will decorate children's rooms. Funny animals a specialty, but will comply with any unreasonable request.

It seems to me we should be on guard against at least the extremes of such situations.

There seem to be problems, then, in our attempts at developing highly general theories. We may, in fact, produce instead of general theories families of individual theories which we might call psychometric methods. Or we may be so vague that the theory is unacceptable. Likewise, there are dangers in working with individual, or specific, theories. A recommended strategy is to attempt theories between these two extremes. The smaller, but not individual, theories may net us much in knowledge gained. These theories will be more easily established and experimented on by individuals. And finally we might hope that several such theories would be compatible for amalgamation with more general theories.

Manuscript received 8/10/55

PSYCHOMETRIKA—VOL. 20, NO. 4
DECEMBER, 1955

F-TEST BIAS FOR EXPERIMENTAL DESIGNS OF THE LATIN SQUARE TYPE

NEIL GOURLAY

UNIVERSITY OF BIRMINGHAM

In an earlier paper, a method of analysis, due to Neyman and now known generally as variance component analysis, was used to examine F-test bias for experimental designs in education of the randomized block type. The same method is now applied to study F-test bias for designs of the Latin square type. The results, in general, disprove the view that, for a valid application of Latin square techniques, it is necessary that all interactions are zero.

In an earlier paper (**5**), a study was made of F-test bias for experimental designs in educational research of the randomized block type. In this paper, a similar study is made of those designs of which the simple Latin square is the prototype. The B-ratio technique, due to Neyman and described in the earlier paper, is again employed.

As McNemar (**8**) points out, the usual textbook statement of the theory underlying the use of the Latin square implies zero interaction between the main effects. McNemar claims that where the assumption of zero interaction is not met, investigators will obtain too many "significant F's"; he then goes on to conclude that, since significant interactions are so common in psychological research, the Latin square is seldom appropriate and that "it is defensible only in those rare cases where one has sound a priori reasons for believing that the interactions are zero."

In the discussion which follows it will be shown that McNemar is by no means altogether correct in his point of view. It would appear that he has failed to realize that in the field of education and psychology, the application of the Latin square design has progressed beyond the usual simple textbook formulation and that some of the later applications show the need for modification of his rather sweeping generalization. In particular, it will be shown that the Latin square can be applied in several cases where the interactions are not zero; also, that in those cases where bias is present, it may well be negative and not positive as McNemar would maintain—a result which can only increase and not diminish the significance of any F-test. As defined in the earlier paper (**5**), an F-test is said to be positively or negatively biased, if, when the null hypothesis being tested is correct, it gives rise to a larger or smaller proportion, respectively, of significant F-ratios than is warranted by the F-distribution.

It will be helpful to the treatment of our problem if we distinguish between the two main types of interaction that can occur in psychology:

> *Type A*—where each individual (or whatever the unit may be—class, grade, etc.) receives only one of the several treatments applied and is represented by only one measurement in the data to be analyzed. In this case, interaction is between main effects, such as treatments, schools, etc.
>
> *Type B*—where repeated measurements are made on the same individual or group. In this case conditions may differ or different treatments may intervene; earlier measurements (treatments, etc.) may affect, i.e., interact with, those that follow.

The position is still further complicated in that Type B interaction may be accompanied by Type A. Also, besides pure interaction effects, the interaction component of any analysis of variance may contain other terms—described as group errors in the previous paper (**5**).

1. *Applications of the Latin Square Involving Type A Interaction Only*

1.1. *An experiment comparing several methods of teaching some school topic*

For simplicity let us take the case of a 3 × 3 square, say:

		Streams		
		u	v	w
	f	A	B	C
Schools	g	B	C	A
	h	C	A	B

i.e., in each school there are three experimental groups which are subjected to the three methods A, B, and C and which can be classified according to some other factor (e.g., streams). (For the benefit of some readers it is to be explained that in many English schools the children in each grade are assigned to classes according to level of ability. The process is known as *streaming*. A three-stream school is one in which there are three classes in each grade representing three levels of ability.) It will be assumed that the numbers in each group are equal—to avoid bias as discussed in the earlier paper (**5**).

Then we may consider two hypotheses: either (*i*) a particular hypothesis —the methods have the same mean effect when an average is taken for each method over the three schools and the three streams; or (*ii*) a general hypothesis—the methods have the same mean effect when an average is taken over the *total* population of schools (of which the three given schools are a random sample) and the three streams.

The first hypothesis is of little interest to the practical investigator. Furthermore, when real interaction is present between methods, schools and streams, the hypothesis cannot validly be tested by the Latin square design and analysis. There are obviously insufficient data: a factorial design is required. But it does not follow, as McNemar maintains, that the Latin square analysis would be positively biased. It might be positively biased in certain cases, but in others, it would be negatively biased, e.g., when real interaction exists only between two of the three classifications (the Latin square design then becomes effectively factorial).

Now let us consider the bias involved in the Latin square analysis when used to test the general hypothesis. It is now necessary to think of the three schools as a random sample from the *total* population of schools (which, as usual, will be taken as being infinite in number). Also the three schools should be assigned at random to the rows of the Latin square.

Then the mean scores for the nine experimental groups might be represented as follows:

$\pi_f+Q_u+T_A+\eta_{uA}+\epsilon_1$	$\pi_f+Q_v+T_B+\eta_{vB}+\epsilon_2$	$\pi_f+Q_w+T_C+\eta_{wC}+\epsilon_3$
$\pi_g+Q_u+T_B+\eta_{uB}+\epsilon_4$	$\pi_g+Q_v+T_C+\eta_{vC}+\epsilon_5$	$\pi_g+Q_w+T_A+\eta_{wA}+\epsilon_6$
$\pi_h+Q_u+T_C+\eta_{uC}+\epsilon_7$	$\pi_h+Q_v+T_A+\eta_{vA}+\epsilon_8$	$\pi_h+Q_w+T_B+\eta_{wB}+\epsilon_9$

(1)

where

(*i*) The general mean over the three methods, the three streams, and the total population of schools has been taken as zero.

(*ii*) The π-values are the main effects of schools, the π-value for each school being the mean for the school taken over the nine method-stream combinations. (It will be assumed that the total population of π-values has zero mean and variance σ_π^2.)

(*iii*) The Q's are the main effects of streams over the three methods and the total population of schools, and $\sum_3 Q = 0$, where $\sum_n F$ is defined to be the sum of all the terms or expressions of type F, and n denotes the total number of such terms.

(*iv*) The T's are the main effects of the methods over the three streams and the total population of schools, and $\sum_3 T = 0$.

(*v*) The nine η's represent the *real* interaction effects of the three methods and the three streams (over the total population of schools) and are such that

$$\begin{aligned} \sum \eta_{uk} &= 0 = \sum \eta_{vk} = \sum \eta_{wk} \qquad (k = A, B, C), \\ \sum \eta_{lA} &= 0 = \sum \eta_{lB} = \sum \eta_{lC} \qquad (l = u, v, w). \end{aligned} \tag{2}$$

(*vi*) The ϵ's include *real* interaction between schools, on the one hand, and method-stream combinations, on the other, and group and sampling error. Each ϵ could in fact be expressed as

$$\epsilon = \zeta + \xi,$$

where ζ is the *real* interaction term for the school and method-stream combination to which the given ϵ belongs, and ξ is a purely random term made up of group error and sampling error.

The infinite population of ϵ's (of ζ's and ξ's) corresponding to any one cell of the Latin square will have zero mean. We shall assume that the nine populations of ϵ-values (one for each cell) have the same variance σ_ϵ^2 (the usual assumption of homogeneity of variance).

The ϵ's of the cells in any one row of the square will be correlated since their ζ-components are correlated. A few words of explanation might be helpful. For each school there are nine ζ-values (one for each method-stream combination) and their sum is zero; this follows from definition (*ii*) above. Also, for each method-stream combination there will be an infinite population of ζ-values (one ζ for each school). Furthermore, these nine populations of ζ-values will be correlated because the sum of the nine ζ-values for each school is zero. Since schools are assigned at random to the rows of the Latin square, it is not very difficult to see that, while the ζ-values which appear in the three cells of any one row will be correlated with one another, they will not be correlated with the ζ-values in the cells of the other two rows.

The correlations (nine in all) between the ϵ-values may not all be the same. This will happen when heterogeneity of correlation exists between the ζ-interaction effects. [A simple example of this type of heterogeneity is discussed more fully in the first paper (**5**).]

Let the correlations in the first row be denoted by ρ_{12} , ρ_{23} , ρ_{13} , where ρ_{12} represents the correlations between the ϵ's in the first and second cells of the row and so on. Let the other six correlations be ρ_{45} , ρ_{56} , ρ_{46} and ρ_{78} , ρ_{89} , ρ_{79} . Also let $R = \sum_9 \rho$.

Now let us derive the E. V.'s (expected values) of the different sums of squares of the variance analysis over the *total* population of schools. The *total* sum of squares is

$$\sum_9 (\pi_f + Q_u + T_A + \eta_{uA} + \epsilon_1)^2 - \frac{1}{9}\{\sum_9 (\pi_f + Q_u + T_A + \eta_{uA} + \epsilon_1)\}^2. \quad (3)$$

By equations above,

$$\sum_9 (\pi_f + Q_u + T_A + \eta_{uA} + \epsilon_1) = 3 \sum_3 \pi + \sum_9 \epsilon. \quad (4)$$

It follows that the E. V. of *total* sum of squares is

$$6\sigma_\pi^2 + 3 \sum_3 Q^2 + 3 \sum_3 T^2 + \sum_9 \eta^2 + 8\sigma_\epsilon^2 - \frac{2R}{9} \sigma_\epsilon^2 . \quad (5)$$

The sum of squares *between methods* is

$$\sum_3 (M_A - M_B)^2, \tag{6}$$

where

$$M_A - M_B = (T_A - T_B) + \tfrac{1}{3}[(\epsilon_1 + \epsilon_6 + \epsilon_8) - (\epsilon_2 + \epsilon_4 + \epsilon_9)], \tag{7}$$

and

$$\text{E. V.}\ (M_A - M_B)^2 = (T_A - T_B)^2 + \tfrac{1}{9}[6\sigma_\epsilon^2 - 2\sigma_\epsilon^2(\rho_{12} + \rho_{46} + \rho_{89})]. \tag{8}$$

Therefore, the E. V. of sum of squares *between methods* is

$$\begin{aligned} &\sum_3 (T_A - T_B)^2 + 2\sigma_\epsilon^2 - \frac{2R}{9}\sigma_\epsilon^2 \\ &\qquad = 3\sum_3 T^2 + 2\sigma_\epsilon^2 - \frac{2R}{9}\sigma_\epsilon^2 \qquad (\text{since } \textstyle\sum_3 T = 0). \end{aligned} \tag{9}$$

Similarly, the sum of squares *between streams* has E. V.

$$3\sum_3 Q^2 + 2\sigma_\epsilon^2 - \frac{2R}{9}\sigma_\epsilon^2. \tag{10}$$

The sum of squares *between schools* is

$$\sum_3 (M_f - M_g)^2, \tag{11}$$

where

$$\begin{aligned} M_f - M_g = (\pi_f - \pi_g) &+ \tfrac{1}{3}[(\eta_{uA} + \eta_{vB} + \eta_{wC}) - (\eta_{uB} + \eta_{vC} + \eta_{wA})] \\ &+ \tfrac{1}{3}[(\epsilon_1 + \epsilon_2 + \epsilon_3) - (\epsilon_4 + \epsilon_5 + \epsilon_6)], \end{aligned} \tag{12}$$

and

$$\begin{aligned} \text{E .V.}\ (M_f - M_g)^2 = 2\sigma_\pi^2 &+ \frac{1}{9}[(\eta_{uA} + \eta_{vB} + \eta_{wC}) - (\eta_{uB} + \eta_{vC} + \eta_{wA})]^2 \\ &+ \frac{2}{3}\sigma_\epsilon^2 + \frac{2\sigma_\epsilon^2}{9}(\rho_{12} + \rho_{23} + \rho_{13} + \rho_{45} + \rho_{56} + \rho_{46}). \end{aligned} \tag{13}$$

It follows, on reduction, that the E. V. of the sum of squares *between schools* is

$$6\sigma_\pi^2 + \frac{1}{3}\sum_3 (\eta_{uA} + \eta_{vB} + \eta_{wC})^2 + 2\sigma_\epsilon^2 + \frac{4R}{9}\sigma_\epsilon^2. \tag{14}$$

The E. V. of the *residual* sum of squares can now be found by subtraction, but it can also be found directly after the manner of the other variance components. (The second procedure provides a check on the algebra in that the E. V. for *total* should equal the sum of the other E. V.'s). By either method the E. V. for *residual* is

$$\frac{1}{3}\sum_3 (\eta_{uA} + \eta_{vC} + \eta_{wB})^2 + 2\sigma_\epsilon^2 - \frac{2R}{9}\sigma_\epsilon^2. \tag{15}$$

It is to be noted that

$$\sum_9 \eta^2 = \frac{1}{3}\sum_3 (\eta_{uA} + \eta_{vB} + \eta_{wC})^2 + \frac{1}{3}\sum_3 (\eta_{uA} + \eta_{vC} + \eta_{wB})^2.$$

This is easily deduced from (2).

We are now in a position to apply the B-ratio technique to examine the bias involved in the F-test *methods v. residual.* The B-ratio for this test is obtained by (i) applying the null hypothesis that the main effects for methods are equal, i.e., $T_A = T_B = T_C$ ($= 0$ since $\sum_3 T = 0$) and (ii) taking the ratio of the E. V. of the methods variance to the E. V. of the residual variance. It will be seen that it has the value

$$\left(1 - \frac{R}{9}\right)\sigma_\epsilon^2 \Big/ \left[\left(1 - \frac{R}{9}\right)\sigma_\epsilon^2 + \frac{1}{6}\sum_3 (\eta_{uA} + \eta_{vC} + \eta_{wB})^2\right], \tag{16}$$

which will normally be less than unity if there is any real interaction between methods and streams. For the 3×3 Latin square, there are essentially two random arrangements:

$$\begin{array}{ccccccc} A & B & C & \text{and} & A & C & B \\ B & C & A & & B & A & C \\ C & A & B & & C & B & A. \end{array}$$

For the second arrangement, the B-ratio is of the form

$$\left(1 - \frac{R'}{9}\right)\sigma_\epsilon^2 \Big/ \left[\left(1 - \frac{R'}{9}\right)\sigma_\epsilon^2 + \frac{1}{6}\sum_3 (\eta_{uA} + \eta_{vB} + \eta_{wC})^2\right].$$

It obviously leads to the same conclusions that apply to the first arrangement.

This suggests that the bias of the F-test will be negative and not positive as McNemar claimed. But another factor to be allowed for is the heterogeneity of correlation between the ϵ-values. As was shown in the previous paper, this heterogeneity produces positive bias. Whether the combined effect of the two types of bias is positive or negative will depend on various factors. But the point to be emphasised is that, for this application of the Latin square, the bias, if not negative, is at least less than that for the factorial type of experiment. If the bias is unimportant in the one case, it must *a fortiori* be unimportant in the other.

Unequal numbers of cases in the cells of the Latin square (whether proportionate or disproportionate with regard to the three classifications, schools, etc.) will also introduce bias into the F-test. This type of bias is discussed in the earlier paper for the case of the replicated (factorial) type of experiment.

1.2. *Alternative designs*

A single Latin square is not a practical design for a methods experiment. It provides too few degrees of freedom for the estimation of the residual or error variance. To obtain increased precision, two courses are available: either (1.2*a*) the replication of the same type of square; or (1.2*b*) the use of two or more different types of square with or without replication. In both cases, each square and each replication of it requires a separate sample of schools.

There would appear to be no special merit in preferring course (1.2*a*) to (1.2*b*) as some investigators have done—under the belief, presumably, that the use of as many different squares as possible is a necessary part of the randomization process. It is quite sufficient for randomization that the schools are allocated at random to the rows of a square [whether the same square throughout as in (1.2*a*) or to different squares as in (1.2*b*)].

1.2*a*. *The replication of the same square*

For simplicity, we will again take the 3×3 square discussed in section 1.1. Let there be n replications, involving, therefore, $3n$ schools in all. Then by an analysis very similar to that of section 1.1, it can be shown that the expected values for the different components of the variance analysis are as in Table 1 (same notation as before).

Two observations can be made: (*i*) By testing *methods* against *residual within schools* instead of *square residual*, a much more precise test is obtained. Furthermore, this F-test cannot be biased (negatively) as a result of any real interaction between methods and streams. It will, however, like the other test, be subject to any positive bias arising from heterogeneity of correlation between the cells in any one row. (*ii*) A test of interaction between methods and streams is provided by the F-tests: *square residual v. residual within schools* and *rows v. residual between schools.*

If interaction were shown to be present, and it were desirable that the methods should be compared for the three streams separately (instead of an average being taken as for the null hypothesis tested above), this could easily be achieved by analyzing the results for each stream separately. The precision of the tests involved would, however, be poor since school differences would be contained in the error variances.

1.2*b*. *Use of two or more types of square with or without replication*

An analysis similar in type to (1.2*a*) can again be carried out. The precise form the analysis takes will, of course, vary with the types of squares selected and the numbers of replications. To save space no such analysis is reproduced here. It might, however, be pointed out that, where 3×3 squares are involved, there are only two possible types of square and with the same number

TABLE 1

Variance			d.f.	Expected Value of Mean Square
Elements of Square	Methods		2	$(1 - R/9)\sigma_e^2 + (3n/2) \sum_3 T^2$
	Streams		2	$(1 - R/9)\sigma_e^2 + (3n/2) \sum_3 Q^2$
	Rows		2	$(1 + 2R/9)\sigma_e^2 + (n/6) \sum_3 (\eta_{uA} + \eta_{vB} + \eta_{wC})^2 + 3\sigma_r^2$
	Square	Residual / Uniqueness*	2	$(1 - R/9)\sigma_e^2 + (n/6) \sum_3 (\eta_{uA} + \eta_{vC} + \eta_{wB})^2$
Residual	Between Schools		$3(n - 1)$	$(1 + 2R/9)\sigma_e^2 + 3\sigma_r^2$
	Within Schools		$6(n - 1)$	$(1 - R/9)\sigma_e^2$

*The term *square uniqueness* was used by Corrigan and Brogden (3) in an analysis very similar to that above. Their type of application will be discussed later.

of replications of each, the analysis bears a certain similarity to that discussed on p. 283 *et seq.* (See also Table 2.) The conclusions to be drawn with regard to bias are the same as for case (1.2*a*).

1.3. *More complicated applications*

Consider the designs given in Figure 1.

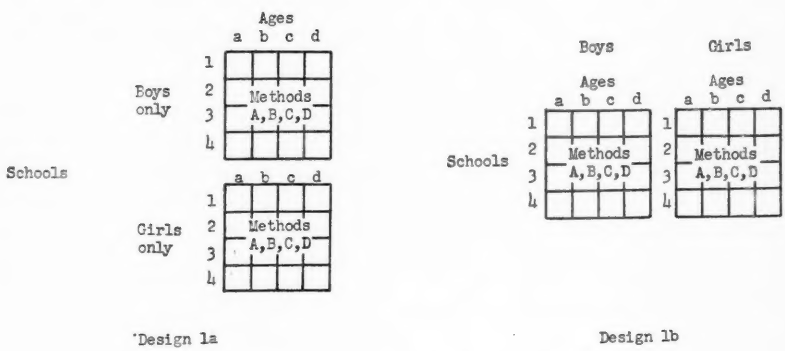

FIGURE 1

Design 1*a*, or something very similar, was used by Burt and Lewis (**2**). Once again it can be shown that the *B*-ratio for design 1*a* is less than unity, although as in all the other cases discussed, positive bias can result from heterogeneity of correlation between the four cells in each of the eight rows.

But the same cannot be said for design 1*b*. In this case it can be shown that with real interaction present, the *B*-ratio is likely to be greater than unity: a more serious degree of positive bias may, therefore, be present in the *F*-test.

2. *Applications of the Latin Square Involving Type B Interaction*

Most of the Latin square studies reported in journals have been of the type where the subjects of the experiments have been subjected to a succession of treatments and tested for each treatment: they could therefore involve Type *B* interaction [cf. Thomson (**11**), Sutherland (**10**), Grant (**6**), Edwards (**4**), and Archer (**1**)].

The Latin square design requires the order of succession of treatments to vary for the individuals and thus makes the problem of interaction more complicated than that which deals with Type *A* interaction only.

2.1. *The case of a single Latin square*

For example, consider three individuals subjected successively to three treatments A, B, and C in the following orders:

	1	A	B	C
Individuals	2	B	C	A
	3	C	A	B

With interaction present (whether type A or B), as stated earlier, the square does not provide sufficient data to test any worth-while hypothesis for the three treatments and the three individuals.

Can a general hypothesis be tested? Is there any difference between treatments when averaged over the total population of individuals of which our three cases are a sample? Once again the data are insufficient if type B interaction is present. The square involves only three of the possible six orders for the treatments, and all six orders would have to be considered in order to obtain a worth-while generalized result. We will examine the latter possibility presently. The only conclusion to be drawn is that a single Latin square is of little use when type B interaction is present or suspected. This, of course, is in accord with McNemar's point of view.

2.2. *Replication of the same square*

In this case each of the treatment sequences is applied not just to one individual but to a group of individuals. The applications of both Thomson (**11**) and Sutherland (**10**) fall into this category. (Sutherland's square is a Greco-Latin square but the same principles apply.) The method was also used by Corrigan and Brogden (**3**), whose application was discussed by Grant (**6**). Edwards (**4**) also illustrates the method.

In outward form, the analysis is very similar to that already discussed on p. 279. *Schools* is replaced by *individuals* and *rows* now represents groups of individuals undergoing different treatment sequences. *Streams* is replaced by some other classification. But there are some important differences which, to save space, we will only indicate:

(*i*) Besides possible type A interaction terms, there may be also type B interaction terms affecting each of the first four variance components (see Table 1).

(*ii*) In the earlier analysis, group errors (i.e., errors, other than random error, peculiar to a school group) were included in σ_ϵ^2 and, therefore, affected all variance components. But in the present case, it will be seen that group errors will vary from cell to cell of the Latin square but will be the same for all the individual measurements in a cell, i.e., group error variance will form part of the first four variance components of the analysis but not of the two residual variances.

What tests may be applied and what is the position with regard to bias? (*i*) An important test is that of ***square residual*** (or ***uniqueness***) against ***residual within individuals***. If significant, this may indicate either type A or B interaction, or group errors, or some combination of the three; the analysis cannot differentiate. With a significant result, there is little point in proceeding further. The test *treatments v. residual within individuals* would have such a limited interpretation that it would be virtually valueless; as a test of a general hypothesis about treatments, the test *treatments v. square residual* would be biased to an unknown extent.

(*ii*) If non-significance is obtained for the test of *square uniqueness*, a further test of zero interaction (and group error) is provided by taking *rows* (*groups* or *sequences*) against *residual between individuals*, provided the groups were random in the first place.

(*iii*) With both these tests non-significant, the other main effects might be tested against *residual within individuals*. But the reader must be warned against following such a test sequence blindly. It is to be remembered that a statistical test cannot *prove* the null hypothesis on which it is based; although the two preliminary tests are non-significant, it may still be the case that interaction (and/or group error) is present. A priori knowledge as to the likelihood of interaction and group error is obviously important. Where past experience would suggest that no interaction or group error is likely to be present, and the two preliminary tests confirm this, the tests of main effects against *residual within individuals* can be made with some safety. But where interaction or group error is known to be likely, little reliance can be placed on the tests of main effects, even though the preliminary tests give a non-significant result. In other words, the given experimental design is almost useless for dealing with this situation.

Corrigan and Brogden's data (**3**) show non-significance for both preliminary tests. Sutherland's data (**10**) show significance for the second test (the groups were not random) and would appear to show significance also for the first test; this would, of course, invalidate his other tests.

Edwards (**4**, p. 325) seems to regard *square residual* and *residual within individuals* as estimates of the same variance but, as we have seen, this can only be the case when interaction and group errors are zero.

2.3. *Analysis involving complete sets of squares*

When type B interaction is present or suspected, it is obvious that all possible treatment sequences must be considered if a generalized result is to be obtained. Grant (**6**) discusses this case.

Once again, for simplicity, we will consider the case of the 3×3 square. We will assume that individuals are assigned at random to the rows. There are then effectively only two Latin squares involved, corresponding to the six possible orders of treatment ABC, BCA, CAB and ACB, BAC, CBA.

For convenience we will give the eighteen cells of the design the numbers 1 through 18, in the order just stated, for the six sequences.

We will consider the case of n replications of the design, i.e., n individuals will be assigned to each row (or sequence). Then each of the n entries in any one of the eighteen cells may be represented as the sum of six terms of the form

$$\pi + Q + T + \eta + \xi + \epsilon, \tag{17}$$

where

(i) The general mean over the eighteen cells and the total population of individuals (assumed infinite) has been taken as zero.

(ii) The π-values ($6n$ in all) are the main effects of individuals averaged over the six sequences. (It will be assumed that the total population of π-values has zero mean and variance σ_π^2.)

(iii) The Q's (Q_u, Q_v, Q_w) are the main effects of columns, averaged over the six sequences and the total population of individuals, and $\sum_3 Q = 0$.

(iv) The T's (T_A, T_B, T_C) are the main effects of treatments averaged over the six sequences and the total population of individuals, and $\sum_3 T = 0$.

(v) The η's (18 in all) represent the joint effect of type A and B interaction for the 18 cells and are such that the six sums of six η-terms corresponding to the three treatments and the three columns are each zero.

(vi) The ξ's (18 in all) represent possible group or cell errors. (It will be assumed that they are random and that the total population of ξ-values has zero mean and variance σ_ξ^2.)

(vii) The ϵ's ($18n$ in all) represent the residuals within cells after all the other effects have been taken out. It will be assumed that the total population of ϵ's for each cell has zero mean and variance σ_ϵ^2. Then the ϵ's of the cells in any one row will be correlated. Let ρ_{12}, ρ_{23}, ρ_{13} represent the correlations for the first row and so on. Also let $R = \sum_{18} \rho$.

The derivation of the E. V.'s of the different sums of squares in the variance analysis is not reproduced here but the results are given in Table 2.

What conclusions are to be drawn from this analysis?

(i) If one or both of the F-tests *residual between cells v. residual within individuals* and *rows* (or *sequences*) *v. residual between individuals* is significant, non-zero interaction and/or group errors is indicated; the analysis cannot differentiate. It would then be invalid to test *treatments* (or *columns*) against *residual within individuals* unless there was other evidence (possibly arising from the design of the experiment) to show that group errors were not present.

If, on the other hand, the first two tests were non-significant, the test of the significance of *treatments* (or *columns*) might safely be made. (The same warning as appears on p. 283 applies here).

(ii) It is always possible to test *treatments* (or *columns*) against *residual between cells*. The B-ratio for this test is never greater than unity. In this

TABLE 2

Variance		d.f.	Expected Value of Mean Square	
Cells	Methods	2	$(1 - R/18)\sigma_\epsilon^2 + n\sigma_\xi^2$	$+ 3n \sum_3 T^2$
	Columns	2	$(1 - R/18)\sigma_\epsilon^2 + n\sigma_\xi^2$	$+ 3n \sum_3 Q^2$
	Rows	5	$(1 + R/9)\sigma_\epsilon^2 + n\sigma_\xi^2 + (n/15) \sum_6 (\eta_1 + \eta_2 + \eta_3)^2$	$+ 3\sigma_\tau^2$
	Resid. between cells	8	$(1 - R/18)\sigma_\epsilon^2 + n\sigma_\xi^2 + (n/8)[\sum_{18} \eta^2 - (\frac{1}{3}) \sum_6 (\eta_1 + \eta_2 + \eta_3)^2]$	
Residual	Between individuals	$6(n - 1)$	$(1 + R/9)\sigma_\epsilon^2$	$+ 3\sigma_\tau^2$
	Within individuals	$12(n - 1)$	$(1 - R/18)\sigma_\epsilon^2$	

respect the present analysis differs from that for the replication of the same square (see previous section), where type B interaction may affect both *treatments* (or *columns*) and *residual between cells* to give a B-ratio (and therefore a bias) of unknown size.

Before concluding this section it might be of interest to mention that type B interaction is similar to the carry-over effect studied by statisticians in animal science [cf. Patterson (**9**) and Lucas (**7**)]. Further, the experimental model which they consider is very much the same as that treated in this section. Their analysis of variance, however, follows quite a different pattern and permits the testing of a wider range of hypotheses. One of their findings is that there is no bias involved in the testing of *unadjusted* direct effects against the error variance. This agrees with conclusion (*i*) above. (It must be noted that group error does not occur in the animal science experiment.)

2.4. *More complex designs.*

No attempt will be made to consider F-test bias for analyses of more complex designs. It should now be apparent that where tests rest on the assumption of zero interaction and group error, the design should provide a test of this assumption. Also, in cases where such a test proves significant (or where the presence of interaction or group error is known to be likely even though unrevealed by any test), the design should furnish tests of main effects, which, although less precise than those which might otherwise have been used, possess B-ratios not exceeding unity.

Archer (**1**) shows himself to be aware of the limitations of the designs he offers in his paper, but considers that the difficulty could be partially overcome by ensuring that the interactions, for which his methods provide no test, are those which the investigator has decided a priori to be unimportant. There is a danger that these a priori decisions may be purely ad hoc assumptions and bear little relation to actual fact.

REFERENCES

1. Archer, E. J. Some Greco-Latin analysis of variance designs for learning studies. *Psychol. Bull.*, 1952, **49**, 521–537.
2. Burt, C. and Lewis, B. Teaching backward readers. *Brit. J. educ. Psychol.*, 1946, **16**, 116–132.
3. Corrigan, R. E. and Brogden, W. J. The trigonometric relationship of precision and angle of linear pursuit-movements. *Amer. J. Psychol.*, 1949, **62**, 90–98.
4. Edwards, A. L. Experimental design in psychological research. New York: Rinehart, 1950.
5. Gourlay, N. *F*-test bias for experimental designs in educational research. *Psychometrika*, 1955, **20**, 227–248.
6. Grant, D. A. The Latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.
7. Lucas, M. L. Bias in estimation of error in change-over trials with dairy cattle. *J. agr. Sci.*, 1951, **41**, 146–148.

8. McNemar, Q. On the use of Latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398–401.
9. Patterson, M. D. The analysis of change-over trials. *J. agr. Sci.*, 1950, **40**, 375–380.
10. Sutherland, J. An investigation into the prognostic value of certain arithmetic tests at the age of eleven plus. *Brit. J. Psychol., Statist. Sect.*, 1952, **5**, 189–196.
11. Thomson, G. H. The use of the Latin square in designing educational experiments. *Brit. J. educ. Psychol.*, 1941, **11**, 135–137.

Manuscript received 6/18/53

Revised manuscript received 8/19/54

PSYCHOMETRIKA—VOL. 20, NO. 4
DECEMBER, 1955

CHARACTERISTICS OF TWO MEASURES OF PROFILE SIMILARITY

CHESTER W. HARRIS

UNIVERSITY OF WISCONSIN

Analogs of Pearson's coefficient of racial likeness and of Mahalanobis' distance measure have been proposed as descriptive statistics for comparing two individuals. This paper shows that two different definitions of "uncorrelated" variables—one associated with an inverse transformation and the other with a principal-axis transformation—give rise to these two descriptive statistics. The effects of putting the data into certain forms, such as equalizing the variances of the variables or equalizing the means of the persons, prior to using either of the two transformations, are discussed.

The interest in measures for assessing similarity (or dissimilarity) of profiles is reflected in such recent summaries and discussions as those of Osgood and Suci (**5**), Gaier and Lee (**3**), Webster (**8**), Cronbach and Gleser (**2**), and Thorndike (**7**). Several of these papers consider the problem of similarity of profile for two individuals, as contrasted with two groups. The latter problem may be formulated as one of discrimination between groups, and, as Cronbach and Gleser point out, two well-known approaches to its solution have been tried. One is the Pearson coefficient of racial likeness and the other the Mahalanobis distance measure, which is known to be related to Fisher's discriminant function. The analogs of these two measures for the problem of comparing two individuals have been suggested in the discussions mentioned above. For one, Osgood and Suci and, independently, Cronbach and Gleser have suggested a measure that is analogous to the Pearson CRL; also, Cronbach and Gleser suggest a Mahalanobis-type measure and compare and contrast it with the former.

The purpose of this paper is to examine these two proposed measures of profile similarity as descriptive statistics. In order to do this, the concept of Euclidean distance will be reviewed, a matric notation developed to describe a distance measure, and then the distinction between these two measures considered as a distinction between two definitions of uncorrelated variables. Finally, the effects of adopting certain forms of the data upon these measures will be outlined. The treatment here follows from well-known principles of matrix algebra, and consequently does not offer any strictly new propositions. However, it does clarify certain characteristics of these two proposed measures and in so doing may assist in describing them as descriptive statistics.

Euclidean Distance

The Euclidean distance between two points in space is a well-defined concept that has been generalized to a space of any size. Providing that the space, of size k, say, has been defined by a rectangular Cartesian system of reference axes, then the square of the distance between any two points in this space is given by the sum of the squares of the differences between paired coordinates of the two points. A rectangular Cartesian system consists of k mutually perpendicular (orthogonal) axes; the pairing of coordinates is done, of course, with respect to these k reference axes. For example, suppose that four persons are located in a space of size two by the following coordinates with respect to a rectangular Cartesian system:

	Person a	Person b	Person c	Person d
Axis 1	4	0	1	3
Axis 2	9	5	8	6

The square of the distance between persons a and b, say, is given by: $(4 - 0)^2 + (9 - 5)^2 = 32$. Since squares are being summed, the result is obviously the same if we compute, instead, $(0 - 4)^2 + (5 - 9)^2 = 32$.

Designate this matrix as X. A method of securing these Euclidean distances is to operate on the matrix $X'X$, where X' is the conventional transpose of X. For these data, $X'X$ is

	a	b	c	d
a	97	45	76	66
b	45	25	40	30
c	76	40	65	51
d	66	30	51	45

The diagonal elements of $X'X$ are simply the sums of the squares of the coordinates for a given person. The off-diagonal elements are the sums of the paired products of the coordinates for the two persons designated by the row and column headings. Thus, for person a the diagonal element is $(4)^2 + (9)^2 = 97$. The element 45, occurring in row b and column a, and in row a and column b as well, is given by $(4)(0) + (9)(5) = 45$. The square of the distance between persons a and b is then given by $97 + 25 - 2(45) = 32$, as before. This in effect merely uses the principle of rewriting a square of a difference between two terms as the sum of the squares of the terms minus twice their cross-product.

The diagonal elements of the matrix $X'X$ give the squares of the lengths of each person vector in the k-space, and the off-diagonal elements give the scalar products of each pair of person vectors in this k-space. A measure of Euclidean distance between persons is thus given by the indicated operation on the matrix $X'X$, when the matrix X describes the several persons with

respect to a rectangular Cartesian system. This operation may be formulated in matric terms. For any pair of persons, i and j, this operation consists of pre-multiplication by a row vector, E, of this form:

$$\begin{array}{cccccccccccc} \multicolumn{12}{c}{\text{Persons}} \\ a & b & \cdots & (i-1) & i & (i+1) & \cdots & (j-1) & j & (j+1) & \cdots & N \\ [0 & 0 & \cdots & 0 & +1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0] \end{array}$$

followed by post-multiplication by the transpose of this vector. For example, the square of the distance between persons a and b is given by

$$[+1 \quad -1 \quad 0 \quad 0] \cdot \begin{bmatrix} 97 & 45 & 76 & 66 \\ 45 & 25 & 40 & 30 \\ 76 & 40 & 65 & 51 \\ 66 & 30 & 51 & 45 \end{bmatrix} \cdot \begin{bmatrix} +1 \\ -1 \\ 0 \\ 0 \end{bmatrix},$$

which is equal to $(97 - 45) - (45 - 25) = 32$, as before. Thus, the square of the distance between any pair of persons is given by a product of matrices that may be written

$$EX'XE' = D^2.$$

This D^2 may be interpreted in more than one way, depending upon how one defines uncorrelated variables. This problem must now be considered.

Uncorrelated Variables

In order to show the nature of this problem, let us define, loosely, an uncorrelated form as

$$TZZ'T' = \text{a diagonal matrix},$$

where Z is a given matrix of data and T is a transformation. To avoid discussing at this point certain problems of the form of the data, let us specify that Z consists of deviation scores that have been systematically reduced so that the variance of each row of Z equals unity. In other words, the data are taken in a form such that ZZ' is the conventional correlation matrix with units in the diagonals. Later, questions concerning the form of the data will be raised and Z will be shown to be a product of matrices, one of which is the matrix of raw scores. Now the definition of an uncorrelated form given above does not specify the non-zero elements in the diagonal matrix; in other words, it does not specify the weighting to be given each of the uncorrelated variables. Two systems of weighting appear to have special merit; one is given by

$$TZZ'T' = I$$

and is loosely related to the Mahalanobis distance measure for groups. The other is given by

$$TZZ'T' = D_\lambda^2 ,$$

where D_λ^2 designates the matrix of non-zero latent (or characteristic) roots of the matrix ZZ'; this definition is associated with the Cronbach-Gleser D^2 and, as they point out, with the Pearson CRL. These two weighting systems give different results; one weights the uncorrelated variables, i.e., the factor scores, equally; the other weights the factor scores in proportion to the size of the square roots of the latent roots.

The Inverse Transformation

First consider the transformation that yields equally weighted uncorrelated variables. It always is possible to resolve Z into a product of principal-axis factors and factor scores; thus

$$Z = GD_\lambda P',$$

where G is a set of orthogonal columns constituting the characteristic vectors (in standard form) corresponding to the non-zero latent roots of ZZ', D_λ is the matrix of positive square roots of the non-zero latent roots of ZZ', and P' is the set of factor scores with unit variance. There now are available these generalities: $G'G = I_r$, where r is the rank of Z, and $P'P = I_r$. GG' is a pre-multiplication unit for Z and consequently a right and left unit for ZZ'; this is true regardless of the rank of Z. Similarly, PP' is a right and left unit for $Z'Z$, and in fact the columns of P are the characteristic vectors of $Z'Z$ corresponding to the non-zero roots of $Z'Z$, which necessarily are the same as the non-zero roots of ZZ'. For a summary, see Harris (**4**).

These properties give a solution for T. Set

$$X = D_\lambda^{-1}G'Z = P'.$$

Then $XX' = I_r$, and the transformation is $T = D_\lambda^{-1}G'$. This principle of transformation gives as the uncorrelated data, X, the principal-axis factor scores of the persons with unit variance. With Z as defined above, these factor scores have means of zero. For the purpose of determining distances between pairs of persons this transformation leads to

$$EX'XE' = EPP'E',$$

that is, the form $X'X$ is simply the form PP'. The effect, then, of this transformation is to give distance measures that are functions of the equally-weighted principal-axis factor scores.

It is conventional to ask concerning the solution of any problem in what sense, if any, the solution is unique. Consider PP'. If $Z'Z$ is non-singular, as it might be, for example, if N, the number of persons, is less than k, the

number of variables, then PP' necessarily is simply the identity matrix, I. This means then that, using this transformation principle, studying relatively few persons with respect to relatively more linearly independent variables always yields the same numerical value for the distance between every pair of persons, regardless of what set of variables was used. If N is greater than k, then $Z'Z$ necessarily is singular and the matrix PP' is a singular idempotent matrix that is a multiplication unit for a group (in the algebraic sense) of singular matrices. This also means that there are many sets of data that will yield the same matrix, PP', when the inverse transformation is made and the resulting $X'X$ calculated. In other words, under these conditions the distances between pairs of persons are not unique to a given set of data. For example, if we pre-multiply any given set of data, Z, by a non-singular matrix we leave invariant the matrix PP', but not, of course, P itself. Since distance measures computed from data that have been transformed by this inverse transformation are functions of PP', this lack of uniqueness to the given data should be recognized. It also should be recognized that these comments assume that the inverse transformation is developed from the data in hand rather than from data for a different group, such as a normative group. If the latter is done, these statements do not hold.

A direct, but quite arduous, calculation procedure would be to factor either ZZ' or $Z'Z$ in order to determine P, the matrix of factor scores. Another calculation method results from the identity

$$PP' = Z'(ZZ')^{-1}Z,$$

provided, of course, that ZZ' is non-singluar. Still another calculation procedure is to utilize the principle of Rao's transformation (6) to develop a triangular matrix, C, such that

$$CZ = X.$$

Then

$$X'X = Z'C'CZ,$$

where $C'C$ is the inverse of ZZ', provided it exists. It is interesting to observe that this latter method works even though ZZ' is singular. Adopting a new notation,

$$C'C = (ZZ')^{-I} = GD_\lambda^{-2}G',$$

and,

$$Z'(ZZ')^{-I}Z = PD_\lambda G'GD_\lambda^{-2}GD_\lambda P' = PP',$$

as before. This analysis uses the principle that if ZZ' is singular, then there exists a matrix $(ZZ')^{-I}$, which also is singular, such that

$$ZZ'(ZZ')^{-I} = (ZZ')^{-I}ZZ' = GG',$$

where GG' is the symmetric idempotent matrix that is a unit for multiplication within the group. The factored form of $(ZZ')^{-I}$ is then seen to be $GD_\lambda^{-2}G'$.

Principal-Axis Transformation

The inverse transformation discussed above gives as the uncorrelated form of the variables a diagonal matrix whose non-zero entries each equal unity. In other words, the inverse transformation gives uncorrelated variables of equal (unit) variance. As noted above, a different transformation may be defined by requiring that the transformation matrix, T, be such that

$$TZZ'T' = D_\lambda^2 ,$$

where, as before, D_λ^2 is the matrix of non-zero latent roots of ZZ'. This is the familiar canonical form of a symmetric matrix; as such, it is a well-known definition of an uncorrelated form. It differs from the inverse transformation in that the transformed variables are now weighted unequally, rather than equally, these unequal weights being given by the square roots of the roots of the characteristic equation of the symmetric matrix ZZ'. If this is chosen as the uncorrelated form, then the transformation is accomplished by setting

$$X = G'Z = D_\lambda P'.$$

It then follows that $XX' = D_\lambda^2$, as required. In order to determine distances between pairs of persons, calculate

$$EX'XE' = EPD_\lambda^2P'E' = EZ'ZE',$$

since GG' is a unit for multiplication, as described above, regardless of the rank of Z. In other words, choosing the canonical form of ZZ' as the uncorrelated form of the variables gives distances between pairs of persons as a function of the entries in $Z'Z$. The calculation procedure obviously requires no comment. For distance measures this solution is unique to the given set of data; this is related to the fact that the canonical form of a symmetric matrix is, under certain rather general conditions, itself unique.

It is apparent that many different diagonal matrices might be chosen as the uncorrelated form of the variables. A choice of a transformation must specify the non-zero elements of this diagonal matrix, i.e., it must specify the weights to be assigned to the variables in uncorrelated form. Two such choices that are meaningfully related to common statistical concepts are the identity matrix, I, associated with the inverse transformation, and the diagonal D_λ^2 , associated with the canonical form of a symmetric matrix. For both these transformations distance measures for pairs of persons are functions of factor scores; using the inverse transformation, the factor scores are weighted equally, whereas using the principal-axis transformation they are weighted unequally.

The Form of the Data

Consider now a matrix of data, Y, that consists of the observed measures, i.e., the raw scores. In order to transform these data into the form of Z,

first write

$$YL = [y],$$

with $[y]$ the matrix of deviation scores. For any Y, which is of order k by N, the matrix L which accomplishes the transformation of raw to deviation scores is

$$\begin{bmatrix} \frac{N-1}{N} & \frac{-1}{N} & \frac{-1}{N} & \cdots & \frac{-1}{N} \\ \frac{-1}{N} & \frac{N-1}{N} & \frac{-1}{N} & \cdots & \frac{-1}{N} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{-1}{N} & \frac{-1}{N} & \frac{-1}{N} & \cdots & \frac{N-1}{N} \end{bmatrix}$$

The matrix L is square, symmetric, of order N, and of rank $(N - 1)$. Direct multiplication verifies that $L = L^2$, i.e., that L is idempotent. The matrix L is a quadratic form with roots of unity and is an example of the type of matrix referred to in Cochran's theorem (1). Further, L is a multiplication unit for any vector consisting of N terms that sum to zero; the row vector E employed earlier is such a vector. Note that what is being done here is to take an operation that is ordinarily considered to be an additive one and to write it as a multiplicative operation; this becomes a useful tool in the analysis of certain relationships among matrices. It is now possible to write

$$SYL = Z,$$

that is, to pre-multiply the deviation scores by the appropriate diagonal matrix to secure the form Z. The diagonal matrix is, of course, one in which each of the non-zero elements is given by the reciprocal of the product of the square root of N and the standard deviation of the variables.

Using this definition of Z and noting that S is a non-singular matrix, that L is idempotent, and that L is a multiplication unit for E, the distance between any pair of persons under the inverse transformation becomes

$$EZ'(ZZ')^{-1}ZE' = EY'(YLY')^{-1}YE'.$$

(A method of calculation when ZZ' is singular was suggested above.) Reduced to these terms, then, this distance measure is a function of the raw scores and of the inverse of the matrix of variances and covariances. An analogous reduction of the Cronbach-Gleser measure gives

$$EZ'ZE' = EY'S^2YE',$$

showing that it is a function of the raw scores and the reciprocals of the variances of the variables, since the scalar $1/N$ affects all pairs in the same way. Clearly, the two measures are identical if, and only if, YLY' is a diagonal matrix of variances. It also is evident that

$$EY'S^2YE' \neq EY'YE',$$

that is, that any change in scale for one or more variables affects the Cronbach-Gleser measure. This is not true for the measure derived from the inverse transformation.

Other modifications of the form of the data might be explored using these techniques. One such modification that is of interest is secured by centering Y by columns rather than by rows; we may write MY to designate this, where M is of the form of L with k substituted for N. The matrix MY is such that each column sums to zero, i.e., the "elevation" has been equalized for all persons. Then by choosing the appropriate diagonal matrix, S_N , the product $S_NY'MYS_N$ yields the intercorrelations of the persons. If the principal-axis factors of this matrix are taken as the descriptions of persons in terms of uncorrelated variables, then the distance between any pair of persons is simply $ES_NY'MYS_NE'$. Obviously,

$$ES_NY'MYS_NE' \neq EY'MYE',$$

which merely states that any change in scale for the columns affects this type of measure.

Finally, MYL is a double-centered matrix, i.e., it sums to zero both by rows and by columns. Since matrix algebra is a linear associative algebra, it makes no difference whether one first forms MY and then centers by rows, or first forms YL and then centers by columns; the resulting MYL is the same in either case. Now, since L is a multiplication unit for any E, it can be seen that

$$EY'MYE' = ELY'MYLE'.$$

This identity emphasizes the point that if the data are centered by columns then *double*-centering does not alter this particular distance measure.

Table 1 gives the algebra of distance measures developed on the basis of the two principles discussed above for various forms of the data. It is evident that not all the distance measures listed there would be judged to be meaningful ones; it also is evident that some of them are duplicates. The table is probably of primary value in demonstrating that this system of analysis makes explicit a range of choices of distance measures and provides a method of specifying the choice that a worker may make. Table 2 gives illustrative scores on three tests for five students. These scores have been used to compute the illustrative distance measures between pairs of students that are given in Table 3.

TABLE 1

Distance Measures for Various Forms of the Data

Form of Data	Principal-axis Transformation	Inverse Transformation
Y	$EY'YE'$	$EY'(YY')^{-1}YE'$
YL	$EY'YE'$	$EY'(YLY')^{-1}YE'$
SYL	$EY'S^2YE'$	$EY'(YLY')^{-1}YE'$
MY	$EY'MYE'$	$EY'M(MYY'M)^{-1}MYE'$
MYS_N	$ES_NY'MYS_NE'$	$ES_NY'M(MYS_N^2Y'M)^{-1}MYS_NE'$
MYL	$EY'MYE'$	$EY'M(MYLY'M)^{-1}MYE'$

TABLE 2

Illustrative Scores on Three Tests for Five Students

	Student				
	a	b	c	d	e
1. Spelling	18	21	11	16	15
2. Usage	44	43	39	51	46
3. Vocabulary	18	17	14	14	20

TABLE 3

Illustrative Distance Measures between Pairs of Students

$EY'YE'$

	a	b	c	d
b	11			
c	90	125		
d	69	98	169	
e	17	54	101	62

$EY'(YY')^{-1}YE'$

	a	b	c	d
b	.32			
c	.52	1.46		
d	1.22	1.54	.73	
e	.47	1.56	.32	1.88

$EY'S^2YE'$

	a	b	c	d
b	1.07			
c	9.03	11.82		
d	6.48	8.08	11.61	
e	1.82	5.52	11.25	8.33

$EY'(YLY')^{-1}YE'$

	a	b	c	d
b	.37			
c	1.23	1.84		
d	1.30	1.80	2.00	
e	.65	1.99	1.92	1.89

REFERENCES

1. Cochran, W. G. The distribution of quadratic forms in a normal system with applications to the analysis of variance. *Proc. Cambridge Phil. Soc.*, 1934, **30**, 178–91.
2. Cronbach, L. J., and Gleser, G. C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–73.
3. Gaier, E. L., and Lee, M. C. Pattern analysis: the configural approach to predictive measurement. *Psychol. Bull.*, 1953, **50**, 140–48.
4. Harris, Chester W. Relations among factors of raw, deviation, and double-centered score matrices. *J. exp. Educ.*, 1953, **22**, 53–58.
5. Osgood, C. E., and Suci, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–62.
6. Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.
7. Thorndike, R. L. Who belongs in the family? *Psychometrika*, 1953, **18**, 267–76.
8. Webster, H. A note on profile similarity. *Psychol. Bull.*, 1952, **49**, 538–39.

Manuscript received 4/19/54
Revised manuscript received 11/8/54

PSYCHOMETRIKA—VOL. 20, NO. 4
DECEMBER, 1955

THE ESTIMATION OF THE DISCRIMINAL DISPERSION IN THE METHOD OF SUCCESSIVE INTERVALS*

RAYMOND H. BURROS

INSTITUTE FOR MOTIVATIONAL RESEARCH, INC., CROTON-ON-HUDSON, N. Y.

A new algebraic formula is derived for estimation of the discriminal dispersion in the method of successive intervals. The legitimate use of the formula requires that as many normal deviates as possible be present in the matrix. For this reason, it is recommended that deviates corresponding to the interval (0.01, 0.99) of the cumulative proportions be used, instead of those corresponding to (0.05, 0.95), the interval used by Edwards and Thurstone. Computations on data published by Edwards and Thurstone showed that when adjustment was made for variability in dispersions calculated by the formula of this paper, a reduction of fifty per cent in mean absolute discrepancy was produced. Since the formula is easy to use and avoids the disadvantages of its predecessors, it should have fairly wide applicability in psychological research.

The method of successive intervals is perhaps the most practical way of obtaining *rational* scale values of stimuli along a unidimensional psychological continuum not simply correlated with any physical variable. The data may be provided by any procedure in which judges classify stimuli into a finite number of mutually exclusive and exhaustive classes which are ordered along some dimension.

When the number of stimuli is small, they may be ranked without ties, so that the number of classes equals the number of stimuli. When the number of stimuli is large, they may be either sorted into piles or rated on a rating scale. With either of these procedures, the number of classes may be considerably less than the number of stimuli. For adequate reliability, a large sample of judges is needed when any of these techniques of gathering data is used.

Although successive intervals was developed by L. L. Thurstone, its first published account was given in a paper by Saffir (**8**) in 1937. Recently, papers by Edwards (**4**) and Edwards and Thurstone (**5**) have presented a

*This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25726 monitored by Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government. The writer is grateful to Dr. A. L. Edwards for a critical reading of an earlier version of this paper, and to Dr. L. H. Lanier and Dr. L. M. Stolurow for editorial advice on the present version, which was written at the University of Illinois. The editors of *Psychometrika* have informed the writer that H. J. A. Rimoldi and M. Hormaeche (**7**) have independently derived the same formula for the discriminal dispersion from a different set of postulates using the law of comparative judgment.

check on internal consistency which indirectly tests the applicability of the postulates to any particular set of data. This check now makes successive intervals a serious rival to the method of paired comparisons. The advantage of successive intervals over paired comparisons lies in its greater speed in collecting data. Empirical studies (**4**, **5**) have shown that there is a linear relation between scale values obtained by these two methods.

In any stimulus scaling method developed in the Thurstone manner, there are at least two important kinds of parameters, represented respectively by S_j , the scale value of the jth stimulus, and σ_j , the corresponding discriminal dispersion. Although adequate computational techniques for estimating each S_j by the method of successive intervals have been published (**4**, **5**), those available for estimating σ_j are subject to improvement.

The first technique, developed by Thurstone and presented by Saffir (**8**), does not base the computation of each σ_j on all of the data. Also, it does not use a simple algebraic formula in the manner originated by Thurstone (**9**, **10**) and further applied by Burros (**2**) and Burros and Gibson (**3**) for estimation of σ_j in the method of paired comparisons. It is interesting to note, therefore, that in a recent paper on successive intervals, Edwards and Thurstone (**5**) did not use the technique presented by Saffir for estimating the dispersions. Instead these writers used one published by Attneave (**1**). A critical examination of Attneave's technique will be made later on in this paper. In the writer's opinion, it does not have a rigorous basis.

Perhaps the most rigorous approach to the problem is a least squares solution recently published by Gulliksen (**6**). Unfortunately, it is possible (although admittedly improbable) that negative estimates of the dispersions may be calculated by this technique. This sort of result could happen if the dispersions are exceedingly variable. A small positive dispersion could then be estimated as negative when his least squares solution is applied to the data. A related discussion of this problem of absurd results in paired comparisons is presented by Burros and Gibson (**3**, pp. 63–64).

Since the techniques published by Saffir (**8**) and Attneave (**1**) are questionable, and the one by Gulliksen (**6**) conceivably may give absurd results, a new formula may be of interest. This paper, therefore, presents the derivation of a simple formula for the estimation of σ_j in the method of successive intervals, which is similar to those previously derived for paired comparisons (**10**, **2**, **3**). The use of the formula will then be illustrated by means of further analysis of data presented by Edwards and Thurstone (**5**).

Definition of Symbols

R $\equiv$ postulated unidimensional psychological continuum with finite range arbitrarily divided into N class intervals corresponding to the steps on an N-point rating scale;

$S \equiv$ postulated unidimensional psychological continuum with unrestricted range corresponding to R;
$R'_k \equiv$ upper true limit of kth class interval of R;
$S'_k \equiv$ corresponding upper true limit of kth class interval of S;
$\dot{R}_j \equiv$ momentary estimate by a judge of the scale position of the jth object on R;
$\dot{S}_j \equiv$ corresponding momentary estimate of the scale position of the jth object on S;
$S_j \equiv$ scale position (mean and median of $\dot{S}_j$) of jth object on S;
$\sigma_j \equiv$ disciminal dispersion or standard deviation of distribution of $\dot{S}_j$;
$X_{jk} \equiv (S'_k - S_j)/\sigma_j$;
$\dot{z}_j \equiv (\dot{S}_j - S_j)/\sigma_j$;
$P_{jk} \equiv$ probability that $\dot{R}_j \leq R'_k$.

Postulates

1. There exists a unidimensional psychological continuum (R) with a finite range, arbitrary units, and an arbitrary origin.
2. There exists a corresponding unidimensional psychological continuum (S) with an unrestricted range, equal units, and an arbitrary origin.
3. $S = f(R)$, where the function is monotonic, increasing, and generally nonlinear.
4. For object j, and corresponding to each observed momentary estimate ($\dot{R}_j$) by a judge on R, there exists a theoretical momentary estimate ($\dot{S}_j$) on S. The distribution of $\dot{S}_j$ is normal with mean (and thus median) of S_j and standard deviation σ_j .

Basic Theorem

Since $f(R)$ is monotonic increasing,

$$P_{jk} \equiv P(\dot{R}_j \leq R'_k) = P(\dot{S}_j \leq S'_k).$$

But if $\dot{S}_j \leq S'_k$, then

$$\dot{S}_j - S_j \leq S'_k - S_j \qquad \text{and} \qquad (\dot{S}_j - S_j)/\sigma_j \leq (S'_k - S_j)/\sigma_j ,$$

so that

$$\dot{z}_j \leq X_{jk}$$

by definition of these quantities. Therefore,

$$P_{jk} = P(\dot{z}_j \leq X_{jk}) = G(X_{jk}),$$

where G is the normal probability integral. Given each estimated value of P_{jk} determined by the empirical frequency distribution of $\dot{R}_j$ on R, therefore, the corresponding estimated value of X_{jk} can be found from a table of the normal integral. These may be arranged in a matrix X.

Whenever P_{jk} equals 0 or 1, the value of X_{jk} is indeterminate. If any proportion is too near to either 0 or 1 (say, less than 0.01, or greater than 0.99) the values of X_{jk} are too unreliable to be recorded. Whenever a value of X_{jk} is indeterminate or unreliable, it is omitted from the X matrix.

Derivation of Formula for σ_j

From the definition of X_{jk}, it follows that

$$S'_k = S_j + \sigma_j X_{jk} . \tag{1}$$

Similarly for X_{ik},

$$S'_k = S_i + \sigma_i X_{ik} . \tag{2}$$

Therefore,

$$S_j + \sigma_j X_{jk} = S_i + \sigma_i X_{ik} \tag{3}$$

and

$$X_{jk} = (S_i - S_j)/\sigma_j + (\sigma_i/\sigma_j)X_{ik} . \tag{4}$$

Equation (4) says that the jth row in the X matrix is theoretically a linear function of the ith row with slope of

$$m = \sigma_i/\sigma_j . \tag{5}$$

Theoretically these two rows are perfectly correlated.

Let V_i and V_j be the respective measures of variability, e.g., standard deviations or ranges, of the ith and jth rows of X. Assuming perfect correlation, therefore, the slope of (4) is also equal to

$$m = V_j/V_i . \tag{6}$$

Therefore,

$$\sigma_i/\sigma_j = V_j/V_i \tag{7}$$

and

$$\sigma_i V_i = \sigma_j V_j . \tag{8}$$

Thus, for any two stimulus objects i and j either side of (8) theoretically equals a constant, defined as

$$\alpha = \sigma_j V_j . \tag{9}$$

Therefore,

$$\sigma_j = \alpha/V_j . \tag{10}$$

In order to estimate α, a unit of σ_i must be chosen. This is an arbitrary matter. The simplest definition is that the unit is the mean of the sigmas of the n stimuli, i.e.,

$$(\sum_i \sigma_i)/n \equiv 1 \tag{11}$$

and

$$\sum_i \sigma_i = n. \tag{12}$$

Summing (10), and using (12),

$$n = \sum_i \sigma_i = \alpha \sum_i (1/V_i). \tag{13}$$

Therefore,

$$\alpha = n/\sum_i (1/V_i). \tag{14}$$

Thus, α is the harmonic mean of the values of V_i, which are obtained empirically from the rows of the X matrix. After α is estimated from (14), each estimated value of σ_i is given by (10).

Now that the new formula has been derived, it is possible to criticize Attneave's technique. According to Attneave, this "assumes that the *mean* dispersion of stimuli represented in one dichotomy is equal to the mean dispersion of those represented in another; this assumption may be only approximately correct" (**1**, p. 340). A sufficient condition for this assumption is that all entries in the X matrix are present. When this is so, it follows from Attneave's directions that the discriminal dispersion of any stimulus equals the ratio of the *arithmetic* mean of the ranges of stimulus X values to the range of the given stimulus. Equations (10) and (14) show, however, that the proper average for the numerator of the ratio is the *harmonic* mean, *not* the *arithmetic* mean. Even when Attneave's assumption is known to be true, therefore, his technique is not strictly correct. It may sometimes give adequate results, however, if the arithmetic mean range and the corresponding harmonic mean are approximately equal.

Application of Formula

To save space, tables presented by Edwards and Thurstone (**5**) will not be reproduced. They provide the following relevant data: (a) their Table 1 (**5**, p. 172) gives the cumulative proportions P_{jk} for ten stimuli rated on a nine-point scale; (b) their Table 2 (**5**, p. 173) gives the normal deviates X_{jk} corresponding to the proportions in the closed interval (0.05, 0.95). In order to reduce the number of empty cells, the writer entered into a copy of this table those additional deviates required to encompass the interval (0.01, 0.99) of the proportions.

Table 1 of this paper shows the results of the computations. V_j is the standard deviation of the reliable normal deviates for stimulus j based upon N_j values. The parameter α is then computed by (14) to be

$$\alpha = n/\sum_j (1/V_j) = 10/8.332 = 1.20.$$

Then each value of σ_j is computed in Table 1 as

$$\sigma_j = \alpha(1/V_j).$$

TABLE 1

Calculation of the Discriminal Dispersion (σ_j) from Data Presented by Edwards and Thurstone (5)

j*	N_j**	V_j	$1/V_j$	σ_j
1	6	1.08	0.926	1.11
2	7	1.26	0.794	0.95
3	6	1.16	0.862	1.03
4	8	1.25	0.800	0.96
5	8	1.24	0.806	0.97
6	6	1.52	0.658	0.79
7	8	1.34	0.746	0.90
8	7	1.48	0.676	0.81
9	8	0.998	1.012	1.21
10	8	0.950	1.052	1.26
Sum			8.332	9.99***

*As rank order (j) of the stimulus increases, the scale value tends to decrease.

**N_j is the number of normal deviates corresponding to proportions in the interval (0.01, 0.99).

***Presumably errors from rounding off decimals account for the departure of this sum from the theoretical value (10.00).

Discussion

The estimates of the dispersions calculated for successive intervals by the formula of this paper correspond roughly to the estimates of the same parameters of the same stimuli computed by Edwards and Thurstone (**5**, p. 177) by means of the method of paired comparisons. They reported that the latter dispersions "showed considerable variation, ranging from a low of .52 for stimulus 6 to a high of 1.32 for stimulus 10." (**5**, p. 177). The corresponding successive intervals dispersions in Table 1 of this paper are 0.79 and 1.26, respectively.

Although Edwards and Thurstone (**5**, p. 177) reported comparable variation in σ_j computed from their successive intervals data by Attneave's technique (**1**), they noted the surprising fact that adjustment for variability of dispersions did not improve the goodness of fit measured by the mean absolute discrepancy of the proportions. Since this does not conform to usual

experience with paired comparisons, some additional computations made by the writer may be of interest.

First, using the normal deviates corresponding to proportions in the interval (0.01, 0.99) and the dispersions previously computed by the new formula, the successive intervals scale values of these stimuli were computed by an algebraic technique. Since there was close correspondence with the scale values reported by Edwards and Thurstone, who used the interval (0.05, 0.95), no details about these computations need be given here.

Then the mean absolute discrepancy was computed to be 0.0135. This is half of the value, 0.027, reported by Edwards and Thurstone when adjustment was made in successive intervals for variability in dispersions calculated by Attneave's technique. It is concluded that the mean absolute discrepancy may be considerably less than that reported by Edwards and Thurstone when correction is made for variability in dispersions calculated by the formula of this paper.

In order to fulfill the requirements of the formula, however, the number of empty positions in the X matrix should be reduced. The use of a wider interval of acceptably reliable proportions, i.e., (0.01, 0.99) instead of (0.05, 0.95) will produce this desired result. The use of this wider interval is, therefore, recommended.

Since the formula presented here avoids the disadvantages of its predecessors but is easy to use, it should have fairly wide applicability in psychological research.

REFERENCES

1. Attneave, F. A method of graded dichotomies for the scaling of judgments. *Psychol. Rev.*, 1949, **56**, 334–340.
2. Burros, R. H. The application of the method of paired comparisons to the study of reaction potential. *Psychol. Rev.*, 1951, **58**, 60–66.
3. Burros, R. H., and Gibson, W. A. A solution for Case III of the law of comparative judgment. *Psychometrika*, 1954, **19**, 57–64.
4. Edwards, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, **36**, 118–122.
5. Edwards, A. L., and Thurstone, L. L. An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 1952, **17**, 169–180.
6. Gulliksen, H. A least squares solution for successive intervals. *American Psychologist*, 1952, **7**, 408 (abstract).
7. Rimoldi, H. J. A., and Hormaeche, M. The law of comparative judgment in the successive intervals and graphic rating scale methods. *Psychometrika*, 1955, **20**, 307–318.
8. Saffir, M. A. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2**, 179–198.
9. Thurstone, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286.
10. Thurstone, L. L. Stimulus dispersions in the method of constant stimuli, *J. exp. Psychol.*, 1932, **15**, 284–297.

Manuscript received 6/25/54

Revised manuscript received 12/2/54

PSYCHOMETRIKA—VOL. 20, NO. 4
DECEMBER, 1955

THE LAW OF COMPARATIVE JUDGMENT IN THE SUCCESSIVE INTERVALS AND GRAPHIC RATING SCALE METHODS*

H. J. A. Rimoldi†

UNIVERSITY OF CHICAGO

AND

M. Hormaeche

UNIVERSITY OF VIRGINIA

The law of comparative judgment is applied to the successive intervals and graphic rating scale methods. A procedure for estimating the modal discriminal process and discriminal dispersion of the stimuli, as well as the value of the boundaries of the intervals on the continuum, is given. From the estimated values it is possible to determine the theoretical proportions and to compare them with the actual experimental proportions. The agreement between these values is an indication of the adequacy of the assumptions made.

The rationale and the system of computations described in the present paper developed from a suggestion offered by L. L. Thurstone in one of his courses at the University of Chicago. He suggested an interpretation of the method of successive intervals based on the assumption that, in the process of indicating preferences, a subject will compare the affective value of each stimulus with the affective value represented by the interval limits on the psychological continuum.

In the present study stimuli were presented using three different procedures:

1. *Method of successive intervals.* The subject was presented with equally spaced intervals having reference to degree of interest in an indicated stimulus. His placement of a check mark in any interval was interpreted as indicating that his interest in that stimulus was greater than that represented by the lower limit of the interval and smaller than the interest represented by its upper limit. A pre-test on approximately 30 subjects demonstrated experimentally that the continuum could be defined unambiguously.

*This article is the first part of a larger study conducted at the *Laboratorio de Psicologia, Facultad de Humanidades y Ciencias,* Montevideo, Uruguay, during the years 1951 and 1952. The authors want to thank Dr. L. V. Jones for his critical comments on the manuscript. The authors have been informed by the editors of *Psychometrika* that R. H. Burros (2) has independently reached the same analytic solution for the computation of stimulus dispersions. Dr. Burros has used a set of assumptions different from the ones stated in the present paper.

†Now at Loyola University, Chicago, Ill.

2. *Multiple category method.* This is a variation of the previous procedure. Subjects were instructed to encircle the word or sign (*Yes, yes,* ?, *no, No*) that best represented their interest in the stimulus.

3. *Graphic rating scale.* Here the subject was asked to state his interest in each stimulus by placing a check mark on a straight line without intervals. The location of the check mark on the continuum was interpreted as indicating that the interest of the subject in the stimulus was greater than that represented by the points on the continuum located to the left of the check mark, and smaller than that represented by the points on the continuum located to the right of the check mark.

In all the presentations it was assumed that the subject compared the value of the stimulus with the value represented by the different points on the continuum.

Determination of L_i

Let S_j ($j = 1, 2, \cdots, j, \cdots, n$) represent the modal discriminal process for the jth stimulus, and σ_j ($j = 1, 2, \cdots, j, \cdots n$) the discriminal dispersion for the jth stimulus. L_i ($i = 1, 2, \cdots, i, \cdots, m - 1$) is the modal discriminal process of the boundary between intervals i and $i + 1$, where there are m successive intervals, and d_i ($i = 1, 2, \cdots, i, \cdots, m - 1$) represents the discriminal dispersion of the ith boundary. d_i is generated in a manner similar to that described by Thurstone (**7**) for the discriminal dispersion of the stimuli.

It will be assumed that the stimuli are normally distributed and that, together with the interval limits, they can be located on the same psychological continuum. Throughout the study it will be assumed that we are dealing with Thurstone's case II (**8**), where several individuals made one judgment each.

The origin of the scale will be defined as

$$\sum_j S_j = 0, \tag{1}$$

and the unit of measurement will be

$$(\sum_j \sigma_j)/n = 1. \tag{2}$$

According to the law of comparative judgment (**8**) and to the previous assumptions it is possible to write

$$L_i - S_j = X_{ij}\sqrt{d_i^2 + \sigma_j^2 - 2r_{ij}\sigma_j d_i}\ , \tag{3}$$

where X_{ij} is the normal deviate corresponding to the proportion of times that stimulus j has been placed in a position less preferred than the point L_i. r_{ij} is the correlation between the subject's judgment of stimulus j and the interval boundary i.

It seems defensible to assume that d_i will be very small when compared

with σ_j and that the more precise the definition of the continuum the smaller the value of d_i. This assumption seems to be corroborated by an unpublished investigation by L. V. Jones at the University of Chicago. The empirical evaluation of d_i, in terms of the unit of measurement of (2), demonstrates the magnitude to be no greater than .05 and generally much smaller. Ignoring the value of d_i, (3) becomes

$$L_i - S_j = X_{ij}\sigma_j . \tag{4}$$

Adding and averaging (4) for all stimuli and keeping L_i constant we have, using (1),

$$(\sum_j X_{ij}\sigma_j)/n = L_i . \tag{5}$$

Determination of the Modal Discriminal Process for Each Stimulus

According to (4) it is possible to have as many S_j values as there are L_i points on the continuum. Keeping S_j constant and adding and averaging for all the L_i values, we obtain

$$(\sum_i L_i - \sigma_j \sum_i X_{ij})/(m - 1) = S_j . \tag{6}$$

Determination of the Discriminal Dispersions

Subtracting (4) for stimuli 1 and 2, we have

$$\begin{array}{l} L_i - S_1 = X_{i1}\sigma_1 \\ L_i - S_2 = X_{i2}\sigma_2 \\ \hline S_2 - S_1 = X_{i1}\sigma_1 - X_{i2}\sigma_2 . \end{array} \tag{7}$$

Equation (7) is similar to the basic equation used in the scaling of mental tests (6) and may be written

$X_{i1} = X_{i2}(\sigma_2/\sigma_1) + (S_2 - S_1)/\sigma_1$, where $(S_2 - S_1)/\sigma_1 = K =$ constant. Thus,

$$X_{i1} = X_{i2}(\sigma_2/\sigma_1) + K. \tag{8}$$

There are as many $X_{i1}, X_{i2}, \cdots, X_{ij}, \cdots, X_{in}$ values as there are i points on the continuum. From these values n standard deviations, V_j, may be computed by

$$V_j = \sqrt{(m - 1) \sum_i X_{ij}^2 - (\sum_i X_{ij})^2} / (m - 1).$$

Performing the necessary operations, (8) can be written as

$$V_1 = V_2(\sigma_2/\sigma_1)$$

and consequently

$$\sigma_2 = c/V_2 , \quad \text{where} \quad V_2\sigma_2 = c = \text{constant}. \tag{9}$$

Changing subscripts of σ and V from 1 to j and adding all the resulting

equations,

$$n = c \sum_i (1/V_i) \qquad \text{since} \qquad \sum_i \sigma_i = n,$$

and accordingly

$$c = n/\sum_i (1/V_i). \tag{10}$$

From (9) and (10) it is readily seen that the value of the disciminal dispersion for any stimulus j is given by

$$\sigma_j = n/[V_j \sum_i (1/V_i)]. \tag{11}$$

If the values of the dispersions are equal for all the stimuli, then, plotting (8), the slope of the line will be unity. If it can be safely assumed that the discriminal dispersons for all the stimuli are equal, then according to (2) their value should be unity. Consequently (5) and (6) can be written as

$$L_i = \sum_j X_{ij}/n, \tag{12}$$

and

$$S_j = (\sum_i L_i - \sum_i X_{ij})/(m - 1). \tag{13}$$

Reproduction of the Original Experimental Proportions

From (4)

$$X_{ij} = (L_i - S_j)/\sigma_j . \tag{14}$$

If the discriminal dispersions are assumed to be equal, then

$$X_{ij} = (L_i - S_j). \tag{15}$$

The X_{ij} values thus obtained should be transformed into proportions and these compared with the original experimental values.

Equations (1) to (15) refer to true values. For the purposes of computation a parallel set of equations can be written using sample values instead of true values.

Experimental Results

Subjects were asked to state their degree of "interest in knowing" certain people. The stimuli were names of men and women who were well known to the experimental group. (See Tables 2, 3, 4, and 5).

TABLE 1

Stimuli	Known Values of σ_j	Values of σ_j from Formula (11)
a	.39	.40
b	1.51	1.53
c	.43	.43
d	1.10	1.06
e	.64	.64
f	1.34	1.38
g	1.60	1.55

TABLE 2

Method of Successive Intervals

(Decimal points omitted.)

		1	2	3	4	5	6	7	8	9
Roosevelt	A*	000	006	000	041	065	176	206	318	188
	T*	005	007	015	027	086	162	186	294	218
Leonardo	A	006	012	024	035	076	206	159	259	224
	T	009	011	020	034	096	163	174	274	218
Hitler	A	076	047	029	059	100	141	147	247	153
	T	070	038	048	059	122	151	134	188	190
Garibaldi	A	006	053	064	171	265	276	076	065	024
	T	034	045	077	112	276	211	143	090	002
Marie Antoinette	A	018	065	100	065	194	147	141	176	094
	T	040	038	056	075	169	209	165	174	074
Mussolini	A	047.	094	094	088	206	159	118	159	035
	T	070	055	078	099	190	210	137	123	038
Joan of Arc	A	012	012	012	053	088	235	182	229	176
	T	010	013	025	039	111	184	185	264	169
Shakespeare	A	006	012	018	006	053	129	212	288	276
	T	007	009	015	025	073	129	155	271	316
San Martín	A	012	018	047	065	200	253	194	188	024
	T	011	021	042	071	196	273	196	161	029
Bécquer	A	012	076	094	059	159	176	147	153	124
	T	038	033	052	069	164	208	165	187	084
Isabella	A	041	065	094	082	312	188	141	041	035
	T	054	055	083	110	222	228	135	095	018
Cleopatra	A	059	035	070	065	176	188	153	159	094
	T	058	044	059	075	162	189	149	172	092
Dante	A	035	024	029	041	129	159	182	276	124
	T	028	025	056	036	129	178	166	226	156
Beethoven	A	000	018	024	024	065	141	159	300	270
	T	009	010	019	028	079	139	156	269	291
Cervantes	A	000	006	024	065	176	312	224	188	006
	T	005	013	027	054	169	272	224	197	039
Napoleon	A	006	012	012	047	076	106	153	294	294
	T	009	010	018	033	072	132	155	262	309
Chopin	A	018	018	047	029	094	182	159	229	224
	T	020	019	031	041	107	160	162	242	218
Michelangelo	A	012	012	012	041	053	165	194	288	224
	T	010	011	021	032	090	152	164	272	248
Columbus	A	018	018	047	047	188	282	188	141	070
	T	017	023	042	067	174	237	192	191	057
Gandhi	A	041	053	041	041	218	141	135	224	106
	T	048	034	052	066	152	188	155	192	113
Ruben Darío	A	024	012	047	082	170	170	188	212	094
	T	022	025	042	060	153	214	182	208	094
Madame Curie	A	006	012	006	035	100	170	188	335	147
	T	006	010	018	033	099	178	196	281	179
Pasteur	A	006	000	000	018	047	176	188	412	153
	T	002	003	008	019	070	159	203	330	206
Sarmiento	A	053	041	065	076	188	200	165	159	053
	T	051	044	064	083	183	208	155	155	057
Bolívar	A	006	012	035	106	229	224	165	170	053
	T	010	018	038	063	183	263	204	182	039
Average Discrepancy (per column)		007	009	012	014	021	029	020	029	019

Standard Deviation = .025 Average Discrepancy (total) = .017

*A = Actual proportions; T = Theoretical proportions.

TABLE 3

Graphic Rating Scale

(Decimal points omitted.)

		1	2	3	4	5	6	7	8	9
Roosevelt	A*	006	006	012	018	147	229	241	129	212
	T*	005	008	013	024	168	195	197	187	203
Leonardo	A	006	012	012	024	229	147	200	200	170
	T	008	011	017	030	189	193	192	176	184
Hitler	A	076	047	029	035	194	135	176	182	123
	T	070	036	041	056	218	154	134	120	171
Garibaldi	A	035	041	041	094	447	194	070	065	012
	T	034	040	060	093	383	219	109	048	014
Marie Antoinette	A	070	053	070	065	300	165	112	094	070
	T	078	047	059	077	279	176	125	088	071
Mussolini	A	118	088	053	112	270	129	112	076	041
	T	129	063	072	088	281	155	101	064	047
Joan of Arc	A	018	024	018	041	265	194	165	147	129
	T	018	020	028	045	237	204	177	144	127
Shakespeare	A	000	012	024	012	141	218	194	176	223
	T	002	003	007	015	147	204	224	211	187
San Martín	A	012	012	041	094	294	212	194	082	059
	T	014	019	031	057	308	245	172	104	050
Bécquer	A	035	053	082	053	259	200	129	094	094
	T	052	037	049	068	274	190	143	103	084
Isabella	A	053	088	065	112	406	141	100	024	012
	T	066	059	081	113	379	179	081	033	009
Cleopatra	A	070	082	053	100	276	129	135	070	082
	T	092	050	061	075	274	167	120	086	075
Dante	A	018	018	024	029	259	188	223	153	088
	T	016	017	028	045	250	219	180	139	106
Beethoven	A	018	012	029	018	100	159	129	223	312
	T	016	012	018	026	143	141	148	166	330
Cervantes	A	006	000	018	041	206	223	259	123	123
	T	006	012	014	031	218	227	208	165	119
Napoleon	A	000	018	012	029	159	123	147	165	347
	T	010	010	014	023	135	145	155	174	334
Chopin	A	029	012	041	024	159	194	147	170	223
	T	026	020	026	039	184	161	154	151	239
Michelangelo	A	018	006	018	018	141	176	223	159	241
	T	013	011	019	029	164	165	166	172	261
Columbus	A	029	018	041	088	388	194	106	065	070
	T	028	030	046	067	317	221	151	092	048
Gandhi	A	041	065	041	076	270	165	100	123	118
	T	057	038	047	064	254	180	139	112	109
Ruben Darío	A	029	041	041	041	265	241	165	100	076
	T	032	030	040	062	284	208	154	112	078
Madame Curie	A	006	012	012	076	212	218	153	159	153
	T	010	012	021	036	219	210	194	160	138
Pasteur	A	006	000	000	047	123	270	206	153	194
	T	007	010	015	027	177	193	189	181	201
Sarmiento	A	059	059	053	070	282	188	100	106	082
	T	068	045	053	073	273	179	133	095	081
Bolívar	A	024	035	041	065	274	206	100	141	112
	T	030	027	038	054	260	201	160	124	106
Average Discrepancy (per column)		006	010	010	014	024	024	028	021	012

Standard Deviation = .022 Average Discrepancy (total) = .016

*A = Actual Value; T = Theoretical Value.

TABLE 4

Multiple Category Method
(Decimal points omitted.)

		1	2	3	4	5
Chopin	A*	053	065	112	376	394
	T*	044	094	111	330	421
Cervantes	A	024	118	170	465	223
	T	025	115	172	464	224
Michelangelo	A	024	070	047	353	506
	T	023	062	086	305	524
Sarmiento	A	070	170	188	406	165
	T	070	175	188	396	111
Napoleon	A	006	076	100	276	541
	T	021	061	084	306	528
Roosevelt	A	024	041	094	365	476
	T	019	078	072	335	496
San Martín	A	035	141	165	465	194
	T	035	136	181	448	200
Mussolini	A	182	153	224	288	153
	T	169	205	166	304	156
Columbus	A	070	188	194	347	200
	T	078	170	181	381	190
Joan of Arc	A	029	112	170	276	412
	T	036	098	124	368	374
Bécquer	A	106	212	123	265	294
	T	125	156	140	308	271
Bolívar	A	047	200	112	424	218
	T	055	151	172	410	212
Hitler	A	182	059	118	294	347
	T	152	126	104	247	371
Beethoven	A	012	076	047	229	635
	T	017	046	066	265	606
Garibaldi	A	100	212	247	365	076
	T	094	240	226	359	081
Average Discrepancy (per column)		008	028	023	031	018

Standard Deviation = .028

Average Discrepancy (total) = .022

*A = Actual proportions; T = Theoretical proportions.

TABLE 5

Paired Comparisons

	Beethoven	Michel-angelo	Napoleon	Roosevelt	Chopin	Joan of Arc	Cervantes	Hitler	Bolívar	Bécquer	San Martín	Columbus	Sarmiento	Garibaldi	Mussolini
Beethoven															
Michelangelo	615*														
	579**														
Napoleon	574	562													
	633	587													
Roosevelt	615	704	527												
	674	618	599												
Chopin	645	527	485	509											
	614	552	508	464											
Joan of Arc	728	645	527	320	527										
	663	610	575	516	540										
Cervantes	769	734	598	444	609	562									
	709	670	666	606	591	564									
Hitler	675	657	728	698	580	580	509								
	722	691	698	640	610	587	528								
Bolívar	716	669	686	769	604	586	538	532							
	729	695	698	648	622	602	548	524							
Bécquer	763	710	692	669	633	627	592	485	538						
	732	698	695	655	633	618	575	556	532						
San Martín	710	698	769	757	698	580	574	521	462	521					
	742	719	742	691	637	625	567	540	512	476					
Columbus	740	722	751	692	716	692	657	598	568	544	550				
	758	736	736	702	670	663	622	602	579	544	575				
Sarmiento	828	769	751	716	728	669	669	568	598	586	615	521			
	779	764	788	752	504	691	652	629	599	556	599	508			
Garibaldi	793	817	817	811	704	769	710	598	710	609	704	633	598		
	810	802	821	794	736	742	712	691	663	618	666	575	575		
Mussolini	763	787	917	840	686	692	698	834	633	568	651	503	586	527	
	810	816	869	846	729	748	716	691	648	591	655	536	532	448	
Average Discrepancy (per column)	040	030	034	068	044	041	036	044	025	019	033	022	042	033	043

Standard Deviation = .053 Average Discrepancy (total) = .037

*Top line indicates actual proportions. **Bottom line indicates theoretical proportions.

The stimuli were presented in four different ways: paired comparisons (15 stimuli), successive intervals (25 stimuli), multiple category method (15 stimuli) and graphic rating scale (25 stimuli). Eight arbitrary interval boundaries were superimposed on the continuum to score the results obtained by using the graphic rating scale method. The 15 names included in the paired comparisons were common to all the other methods. Instructions to the subjects stated clearly that the continuum varied from "extreme lack of interest" to "extreme interest" in knowing the persons indicated by the stimuli.

The experimental population consisted of 170 adults, of both sexes, most of them enrolled in teacher training institutions in Montevideo, Uruguay. The tests were administered to groups of 20 to 30 subjects. To check the accuracy of formula (11) a fictitious example with known σ_j values was prepared. The values obtained by using formula (11) and the real values of σ_j are compared in Table 1.

The following operations were performed: *a*) Frequencies and corresponding proportions were obtained for all the cells, (the actual values for these proportions are given in the upper portion of the cells in Tables 2, 3, 4 and 5). *b*) Cumulative proportions were calculated and the corresponding normal deviates determined. *c*) The L_i and S_j values were computed using formulas (12) and (13), Tables 6 and 7. *d*) The values of σ_j were determined by means of formula (11), Table 6, as follows: *i*) from the X_{ij} values the V_j values were computed, *ii*) the sum of all $(1/V_j)$ values was determined, *iii*) noting that in formula (11) $\sum_j (1/V_j)$ and n are constants, for a stimulus j the value of σ_j was obtained by finding first the value of V_j and then applying formula (11). *e*) Improved estimates of L_i and S_j were obtained by means of formulas (5) and (6), Tables 6 and 7. *f*) The original proportions were reproduced using formula (14). (These values are given in the lower portions of the cells in Tables 2, 3, 4 and 5). The paired comparison data were analyzed using Thurstone's cases III and V, as described by J. P. Guilford (**3**).

Figure 1 indicates that the relationship between the S_j values obtained by means of the different procedures here described may be interpreted as linear. The slope of the best-fitting line is an indication of the relative dispersion of the S_j values in the two procedures. The paired comparison procedure gives the maximum scatter while the method of successive intervals reduces this scatter to a minimum. This may be due to the actual manner of presenting the stimuli. It should also be remembered that the best determination of the S_j values by using the paired comparisons method implies an estimation of the correlations between stimuli. As it will be shown in a future paper there are reasons to believe that some of the stimuli used in this study have substantial correlations among themselves. Consequently, this may explain some of the discrepancies found in this study. Our results are in agreement with those reported by Hevner (**4**), Saffir (**5**) and Edwards (**1**).

TABLE 6

	Values of S_j (Formula 13)				Values of S_j (Formula 6)				Values of σ_j			
	Method of Successive Intervals	Graphic Rating Scale	Multiple Category Method	Paired Comparisons (Case V)	Method of Successive Intervals	Graphic Rating Scale	Multiple Category Method	Paired Comparisons (Case III)	Method of Successive Intervals	Graphic Rating Scale	Multiple Category Method	
Roosevelt	.476	.502	.475	.363	.422	.412	.465	.256	.960	.917	1.000	Roosevelt
Leonardo	.342	.384			.369	.329			1.034	.945		Leonardo
Hitler	-.207	-.142	-.238	-.064	-.038	-.038	-.038	-.065	1.377	1.278	1.532	Hitler
Garibaldi	-.601	-.514	-.584	-.704	-.624	-.522	-.605	-.786	.797	.771	.764	Garibaldi
Marie Antoinette	-.296	-.411			-.298	-.406			1.016	1.074		Marie Antoinette
Mussolini	-.591	-.662	-.592	-.712	-.600	-.680	-.600	-.597	.998	1.109	1.054	Mussolini
Joan of Arc	.242	.106	.179	.248	.223	.080	.149	.214	.988	.966	.976	Joan of Arc
Shakespeare	.521	.500			.641	.457			1.107	.963		Shakespeare
San Martín	-.178	-.058	-.060	-.196	-.297	-.154	-.195	-.175	.778	.806	.782	San Martín
Bécquer	-.236	-.255	-.300	-.220	-.234	-.250	-.252	-.265	1.024	1.031	1.179	Bécquer
Isabella	-.658	-.760			-.668	-.720			.876	.804		Isabella
Cleopatra	-.349	-.446			-.318	-.440			1.120	1.122		Cleopatra
Dante	-.013	.097			.054	.037			1.114	.911		Dante
Beethoven	.441	.456	.633	.745	.563	.651	.765	1.045	1.116	1.201	1.112	Beethoven
Cervantes	.034	.330	.062	.052	-.151	.188	-.112	.013	.752	.840	.764	Cervantes
Napoleon	.458	.548	.486	.564	.602	.692	.541	.431	1.134	1.137	1.060	Napoleon
Chopin	.162	.214	.186	.325	.279	.342	.245	.393	1.149	1.177	1.085	Chopin
Michelangelo	.371	.411	.464	.604	.446	.484	.535	.629	1.080	1.081	1.074	Michelangelo
Columbus	-.133	-.230	-.280	-.391	-.208	-.273	-.325	-.459	.878	.871	.903	Columbus
Gandhi	-.230	-.217			-.184	-.186			1.121	1.110		Gandhi
Ruben Darío	-.086	-.149			-.111	-.172			.973	.951		Ruben Darío
Madame Curie	.363	.263			.302	.188			.950	.911		Madame Curie
Pasteur	.675	.436			.471	.384			.855	.950		Pasteur
Sarmiento	-.413	-.343	-.284	-.465	-.419	-.332	-.346	-.492	1.011	1.079	.852	Sarmiento
Bolívar	-.094	-.062	-.148	-.148	-.223	-.069	-.224	-.142	.794	.994	.864	Bolívar
Σ	.000	-.002	-.001	.001	-.001	.002	.003	.000	25.002	24.999	15.001	Σ

TABLE 7

	Values of L_i (Formula 12)			Size of Interval (from Formula 12)		
	Method Successive Intervals	Graphic Rating Scale	Multiple Category Method	Method Successive Intervals	Graphic Rating Scale	Multiple Category Method
L_1	-2.095	-1.955	-1.637	.332	.296	.706
L_2	-1.763	-1.659	-.931	.325	.268	.458
L_3	-1.438	-1.391	-.473	.316	.287	.985
L_4	-1.122	-1.104	.512	.521	.817	
L_5	-.601	-.287		.555	.516	
L_6	-.046	.229		.472	.467	
L_7	.426	.696		.788	.511	
L_8	1.214	1.207				
Σ =	-5.425	-4.264	-2.529			

	Value of L_i (Formula 5)			Size of Interval (Formula 5)		
	Method Successive Intervals	Graphic Rating Scale	Multiple Category Method	Method Successive Intervals	Graphic Rating Scale	Multiple Category Method
L_1	-2.076	-1.936	-1.611	.332	.295	.676
L_2	-1.744	-1.641	-.935	.318	.266	.445
L_3	-1.426	-1.375	-.490	.304	.276	.956
L_4	-1.122	-1.099	.466	.507	.800	
L_5	-.615	-.299		.541	.508	
L_6	-.074	.209		.466	.459	
L_7	.392	.668		.783	.509	
L_8	1.175	1.177				
Σ =	-5.490	-4.296	-2.570			

TABLE 8

Method of Successive Intervals

	1	2	3	4	5	6	7	8
Roosevelt	(-2.478)	(-2.240)	(-1.935)	-1.675	-1.216	-.559	-.015	.885
Leonardo	(-2.335)	-2.097	-1.728	-1.426	-1.024	-.361	.045	.762
Hitler	-1.432	-1.160	-1.028	-.803	-.493	-.121	.251	1.019
Garibaldi	(-1.801)	-1.563	-1.160	-.542	.148	.974	1.347	1.977
Marie Antoinette	-2.097	-1.385	-.904	-.681	-.146	.225	.613	1.316
Mussolini	-1.675	-1.076	-.722	-.459	.073	.490	.863	1.812
Joan of Arc	-2.257	-1.977	-1.799	-1.347	-.927	-.222	.238	.927
Shakespeare	(-2.335)	-2.097	-1.799	-1.728	-1.311	-.739	-.161	.595
San Martín	-2.257	-1.881	-1.426	-1.071	-.407	.240	.803	1.995
Bécquer	-2.257	-1.353	-.908	-.703	-.253	.192	.592	1.155
Isabella	-1.739	-1.248	-.842	-.577	.238	779	1.426	1.799
Cleopatra	-1.563	-1.316	-.987	-.742	-.240	.235	.662	1.311
Dante	-1.812	-1.563	-1.353	-1.131	-.650	-.210	.251	1.150
Beethoven	(-2.335)	-2.097	-1.728	-1.506	-1.122	-.607	-.174	.616
Cervantes	(-2.424)	(-2.186)	-1.881	-1.311	-.610	.210	.867	(1.637)
Napoleon	(-2.335)	-2.097	-1.881	-1.426	-1.024	-.646	-.222	.542
Chopin	-2.097	-1.799	-1.385	-1.216	-.820	-.284	.118	.759
Michelangelo	-2.257	-1.977	-1.799	-1.426	-1.126	-.539	-.028	.762
Columbus	-2.097	-1.799	-1.385	-1.126	-.473	.253	.800	1.468
Gandhi	-1.739	-1.316	-1.103	-.931	-.269	.088	.440	1.248
Ruben Darío	-1.977	-1.799	-1.385	-.974	-.426	.012	.504	1.311
Madame Curie	(-2.335)	-2.097	-1.977	-1.563	-.999	-.443	.043	1.045
Pasteur	(-2.780)	(-2.542)	(-2.237)	-1.977	-1.468	-.684	-.164	1.024
Sarmiento	-1.616	-1.316	-.999	-.722	-.194	.313	.800	1.616
Bolívar	(-2.335)	-2.097	-1.616	-.999	-.284	.284	.762	1.616
n	15	22	23	25	25	25	25	24
Σ	-28.872	-37.110	-31.786	-28.062	-15.023	-1.140	10.661	28.710
Average	-1.925	-1.687	-1.382	-1.122	-.601	-.046	.426	1.196
Dif. Successive Columns	.238	.305	.260	.521	.555	.472	.770	
Σ Total Columns	-52.365	-44.078	-35.958	-28.062	-15.023	-1.140	10.661	30.347
Average (L_i)	-2.095	-1.763	-1.438	-1.122	-.601	-.046	.426	1.214
Size Intervals	.332	.325	.316	.521	.555	.472	.788	

$$\sum_{i=1}^{m-1} L_i = -5.425$$

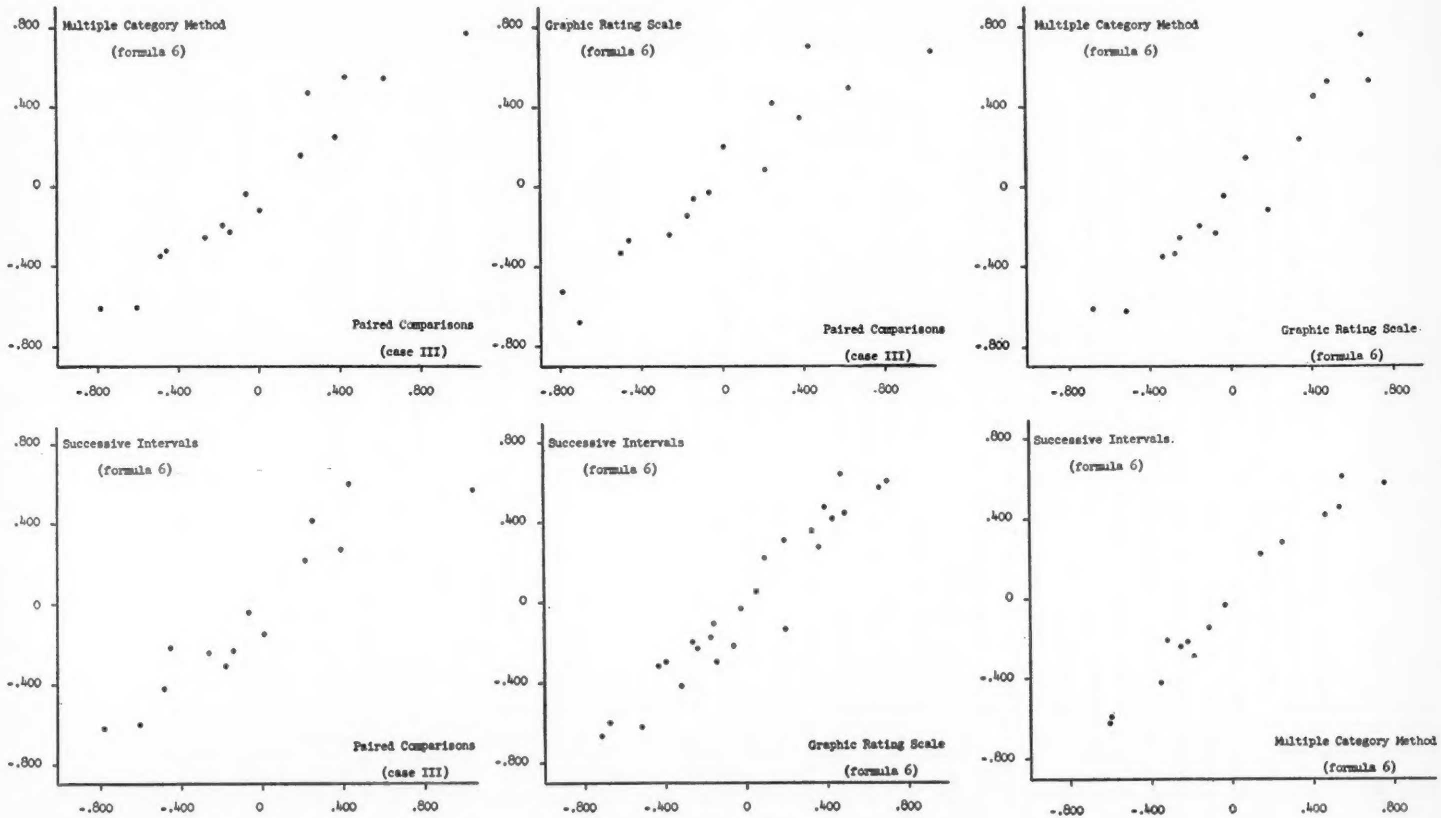
Multiple Category Method
(formula 6)
Paired Comparisons
(case III)
Graphic Rating Scale
(formula 6)
Paired Comparisons
(case III)
Multiple Category Method
(formula 6)
Graphic Rating Scale
(formula 6)
Successive Intervals
(formula 6)
Paired Comparisons
(case III)
Successive Intervals
(formula 6)
Graphic Rating Scale
(formula 6)
Successive Intervals.
(formula 6)
Multiple Category Method
(formula 6)
.800
.400
0
-.400
-.800

FIGURE 1

The procedure employed to deal with extreme X values is as follows: *a*) Cells with cumulative proportions above .990 or below .010 were deleted. *b*) The X values for the remaining cells were calculated and the average for each column was found (Table 8). *c*) In Table 8 the difference between the averages for successive columns was found; for instance, the difference between columns 3 and 4 was .260. *d*) This average was used to determine the estimated value corresponding to the deleted cells in column 3; for instance, for the stimulus Pasteur the new value was $-1.977 + (-.260) = -2.237$, and for Roosevelt it was -1.935. A similar procedure was followed to complete all the missing cells in the table.

The estimated values in Table 8 are given in parentheses (16 values out of 225 for Table 8, 10 out of 225 for graphic rating scale, and 1 out of 75 for multiple category method). As a final check on the operations remember that $\sum S_i = 0$, and $\sum \sigma_i = n$.

The theoretical proportions obtained using formula (14) should agree closely with the actual proportions provided the assumptions used in the development of the method are substantially correct. At the bottom of Tables 2, 3, 4 and 5 the standard deviation and average discrepancies between the actual and theoretical proportions are given. It is clear that the S_i and σ_i values obtained in the manner indicated in this article reproduce the experimental values satisfactorily. Notice that the reproduction of the experimental values using the method here described is better than the one obtained using the paired comparisons method.

REFERENCES

1. Edwards, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, **36**, 118–122.
2. Burros, R. H. The estimation of the discriminal dispersion in the method of successive intervals. *Psychometrika*, 1955, **20**, 299–305.
3. Guilford, J. P. Psychometric methods. New York: Mc-Graw Hill, 1936.
4. Hevner, H. An empirical study of three psychophysical methods. *J. gen. Psychol.*, 1930, **4**, 191–212.
5. Saffir, M. A. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2**, 179–198.
6. Thurstone, L. L. A method of scaling psychological and educational tests. *J. educ. Psychol.*, 1925, **16**, 433–451.
7. Thurstone, L. L. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368–389.
8. Thurstone, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286.

Manuscript received 1/13/54

Revised manuscript received 1/23/55

PSYCHOMETRIKA—VOL. 20, NO. 4
DECEMBER, 1955

A STATISTICAL MODEL FOR RELATIONAL ANALYSIS*

R. DUNCAN LUCE† AND JOSIAH MACY, JR.‡
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
AND
RENATO TAGIURI
HARVARD UNIVERSITY

The diadic relationships existing in a group can be defined in terms of the members' choices, rejections, and their perceptions of being chosen and rejected. The number of possible distinct diads is 45. Formulas are given for computing the expected frequency and variance of the different diadic forms expected, when certain random factors are taken into account. These values must be known if the operation of factors other than the specified random ones is to be studied. Values obtained from two models with different assumptions are compared with empirical values. A simplified treatment is possible for groups with ten or more members.

The student of interpersonal processes often needs to describe and classify in some useful form the relationships between individuals. One such classification is given by relational analysis (2), a method developed in conjunction with a series of studies in interpersonal perception. In this classification the relationship between two persons is described in terms of the *feeling* each has for the other, and the *perception* each has of the other's feeling. More specifically, each member of a well-acquainted group is asked to select those he likes most and those he likes least, and also to guess who likes him most and least. This procedure yields a simple but useful description of the relationship existing between each of the $N(N - 1)/2$ pairs in the group.

Since each subject S_i can choose, reject, or omit any other subject S_j , and can feel chosen, rejected, or omitted by him, nine arrangements are possible of S_i's feelings and perceptions regarding S_j . We shall define a diad between S_i and S_j as any one of the nine possible arrangements of selections of S_i , combined with any one of the nine possible arrangements of selections of S_j , without regard to order. The number of possible distinct diads is 45.

If S_i's feeling toward S_j be denoted by 0 for like, 1 for omission and 2 for dislike and, if S_i's predictions of S_j's feeling toward him be denoted by

*The present problem emerged from research undertaken as part of a project in interpersonal perception being carried out at the Laboratory of Social Relations at Harvard with the financial aid of the Office of Naval Research (Task Order N5ori—07646).

†Now at Columbia University.

‡Now at Johns Hopkins University.

0 for like, 1 for omission and 2 for dislike, then we can represent some of the 45 possible relationships as follows:

S_i	S_j	S_i	S_j
(11)	(11)	(00)	(00)
(01)	(11)	(00)	(22)
(00)	(01)	(22)	(22)

Legend: the first digit in each bracket corresponds to the feeling, the second to the perception.

Some of the possible diads are well integrated, positive, and realistic; others involve contrary feelings and mistaken perceptions; still others indicate a well-developed negative, and recognized, mutual orientation.

Psychologically important features of a group can be described in terms of the frequency of occurrence of the various diads. It is apparent, however, that given the number of choices, omissions, and rejections, and the number of perceptions of choice, omission, and rejection made by each member of a given group, each diad may be expected to occur a certain number of times by chance alone. To interpret observed data we must know something of these chance distributions, so that we will not attempt to give a psychological interpretation to data which can be explained by the operation of chance alone. When we know which specific diads occur from group to group with greater or less than chance frequency, then we can formulate hypotheses about the possible non-chance factors at work. For these reasons it is important to be able to state the expected frequency of occurrence and the variance of each diad type in a group of given size for an assumed chance model. In previous work (3) estimates of these quantities were obtained by constructing a Monte Carlo robot "group," which was set to match the real group man by man in the number of choices, omissions, and rejections made and in their respective perceptions. This is clearly an unsatisfactory and inefficient method if it can be replaced by a simple mathematical formula.

The purpose of this paper is to present a model in terms of which we can estimate the expected chance frequency and variance of the various diads. It should be borne in mind that the distribution of such frequencies is, probably, often more Poisson than normal.

I. *The Model*

Several possible "chance" models are conceivable, depending on what we choose to regard as chance. The first one we shall examine corresponds to the case in which the members of a group are regarded as automata, randomly allocating their selections according to fixed probabilities of choosing, rejecting, or omitting every member of the group. Three other

assumptions are made. First, statistical independence is assumed among the different choices, and between the choices and guesses, made by any individual. Second, the choices and guesses of any subject are assumed to be independent of those made by any other subject. Finally, we assume each subject may not choose or guess the same other subject more than once.

For this model, in other words, we assume that the chance occurrence does not include the operation of any psychological factors except those which govern the relative frequencies of the choices and perceptions. In section III we shall discuss a modification of this, in which we assume an S's perceptions to be conditioned by his choices and rejections.

Let us now proceed with the derivation of the expressions for the expected frequency and variance of each of the diad types. Let S_i's feeling toward S_j be denoted by 0 for like, 1 for omission, and 2 for dislike. Let S_i's prediction of S_j's feeling toward him be denoted by 0 for like, 1 for omission, and 2 for dislike. S_i's statement of his relationship with S_j will be written $(k_1k_2)_{ij}$. Then a diad may be denoted $(k_1k_2)_{ij}(k_1'k_2')_{ji}$, where $k_1 = 0$, 1 or 2 etc., and since we do not consider the order, the diads $(k_1k_2)_{ij}(k_1'k_2')_{ji}$ and $(k_1'k_2')_{ji}(k_1k_2)_{ij}$ are identical. We will sometimes distinguish between them for computational reasons, but in general we shall denote either of these diads by $(k_1k_2)(k_1'k_2')$.

For this model we have assumed that each S_i has a fixed probability $P_i(k_1)$ of liking, not mentioning, or disliking each (other) S_j and that this $P_i(k_1)$ is independent of j; similarly S_i has a fixed probability $Q_i(k_2')$ of predicting these feelings on the part of each S_j, and this is independent of j and of P_i.

Now let X_{ij} be a random variable which assumes the value 1 if the diad between i and j has some specified value $(k_1k_2)(k_1'k_2')$, and is 0 otherwise. Then

$$X = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} X_{ij} \qquad (i \neq j)$$

is a random variable representing the frequency of occurrence of this specified diad in the group. (The following formulas are readily generalized to situations in which any fixed number of categories of questions are answered by the S's and for which the number of possible responses in each category need only be required to be finite. However, many more than two categories with three or four responses each are not very practical.) Since X_{ij} are all independent,

$$E(X) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} E(X_{ij}) \qquad (i \neq j)$$

and

$$E(X_{ij}) = P_i(k_1)Q_i(k_2)P_j(k_1')Q_j(k_2'),$$

so

$$E(X) = \tfrac{1}{2}\{[\sum_i P_i(k_1)Q_i(k_2)][\sum_j P_j(k_1')Q_j(k_2')] - \sum_i P_i(k_1)Q_i(k_2)P_i(k_1')Q_i(k_2')\}.$$

And similarly

$$\operatorname{var}(X) = \tfrac{1}{2} \sum_i \sum_j \operatorname{var}(X_{ij}) \qquad (i \neq j)$$

and

$$\operatorname{var}(X_{ij}) = E(X_{ij})[1 - E(X_{ij})],$$

so

$$\operatorname{var}(X) = E(X) - \tfrac{1}{2}\{[\sum_i P_i^2(k_1)Q_i^2(k_2)][\sum_i P_i^2(k_1')Q_i^2(k_2')] - \sum_i [P_i(k_1)Q_i(k_2)P_i(k_1')Q_i(k_2')]^2\}.$$

Table 1 shows the observed number of choices, omissions, and rejections, and the number of perceptions of choice, omission, and rejection given by each of the members of a ten-man group. The data used as an example were

TABLE 1

Observed Frequencies of Different Feelings and Perceptions in a Ten-Man Group

Subject	Feeling			Perception		
i	k = 0	k = 1	k = 2	k' = 0	k' = 1	k' = 2
1	3	5	1	3	5	1
2	3	2	4	5	1	3
3	6	0	3	4	3	2
4	2	6	1	2	5	2
5	2	4	3	3	3	3
6	4	3	2	1	5	3
7	4	4	1	2	6	1
8	3	4	2	3	4	2
9	2	5	2	2	5	2
10	3	5	1	1	7	1

TABLE 2

Observed and Expected Frequencies of Congruous and Non-Congruous Diads; Model with Perception Not Contingent upon Feeling

Diad Type	Observed	Expected
Bilateral Congruency	20	6.87
Unilateral Congruency	22	21.37
No Congruency	3	16.76
Chi square = 36.32 d.f. = 2 $p < 0.001$		

TABLE 3

Conditional Probabilities $Q(k_2 | k_1)$ for the Ten-Man Group

k_1	$Q(0 \| k_1)$	$Q(1 \| k_1)$	$Q(2 \| k_1)$
0	0.63	0.35	0.02
1	0.24	0.67	0.10
2	0.07	0.40	0.53

TABLE 4

Observed and Expected Frequencies of Congruous and Non-Congruous Diads; Model with Perception Contingent upon Feeling

Diad Type	Observed	Expected
Bilateral Congruency	20	17.63
Unilateral Congruency	22	21.56
No Congruency	3	6.07
Chi square = 1.88 d.f. = 2 $p \doteq 0.40$		

obtained at the end of the last meeting of a series of twelve sessions conducted by a psychoanalyst. The nature of the meetings was a modified form

of group therapy where the members met to "discuss principles of group psychology particularly as these relate to self understanding." The procedure consisted of asking the members to indicate those others in the group they "liked most" and "least" as well as to guess who would name them as liking them "most" and "least."

Analysis of the data in terms of the particular composition of each of the $N(N - 1)/2$ diads gives the observed frequency of each diad type and, in terms of these, describes the group. For example, the diad (00)(00) has an observed frequency of six, while (02)(20) does not occur; the figures below show that these frequencies are different from the expected value predicted by the chance model.

Diad	*Observed*	*Expected*	*Variance*
(00)(00)	6	0.48	0.46
(02)(20)	0	0.50	0.50

The first diad, in which both subjects like each other and predict being liked by the other, occurs more often than expected by chance; the other, in which feelings are not mutual but are accurately predicted, occurs less often than expected, but not significantly so.

In general we have found that there is a significant discrepancy between observed frequencies and those predicted by this chance model. This indicates that there are factors operating other than those we have assumed in this model, and the differences do suggest the nature of some of these factors.

Let us further exemplify the use of the model. It is quite apparent that, in general, the feeling we hold for a person is *congruent* with the feeling we *perceive* that person holds for us. This tendency is quite apparent in all of our data. Thus member S_i tends to choose and feel chosen by S_j, or to dislike and feel disliked by S_j, etc. Is this tendency sufficiently consistent that diads containing such congruencies between feelings and perceptions would exceed chance, while others would fall below chance? The present model permitted us to test such hypotheses by supplying us with an acceptable chance baseline. The data for the ten-man group mentioned above will be used to illustrate this point. We will separate the diads into three groups. In the first we will put all diads in which feeling and perception are the same for both members: (00)(00), (00)(11), (00)(22), etc. In the second, we will put those diads in which this is true for only one member: (00)(01), (00)(12), etc.; in the third we will put all diads in which this holds for neither member: (01)(12), (10)(21), etc. If our conjecture is right, the first class should contain more cases than expected, and the third class fewer than expected, on the basis of chance alone. The figures presented in Table 2 show that the differences are as predicted; the probability of this occurring with the chance model is less than 0.001.

II. *Model with Perceptions Dependent upon the Subject's Feelings*

We may now modify the basic chance model by incorporating various hypotheses about the group, to see whether these additional factors will explain the observed results.

We have said above that, in all groups, we have observed a strong tendency for a member to predict that others feel toward him whatever he feels toward them, and we have exhibited this tendency in one group. It seems reasonable to ask whether this tendency alone accounts for the deviation from chance. We shall, therefore, investigate how well a model with this modification accounts for the data.

We shall assume again that each S responds in an independent manner, with probability $P_i(k)$ of liking, omitting, or disliking any other subject. However, we shall now assume that this choice conditions his prediction of another's feeling toward him, so that his probability of predicting a given response on the part of another member is not $Q_i(k')$, as before, but is $Q_i(k' \mid k)$. The expressions for the expected value for the occurrence of any diad and the variance take similar forms to those presented above, with this conditional probability used for Q_i . In the case of the group used here as an example, we do not have sufficient data to estimate the conditional probabilities individually for each member, so we shall use one set of such probabilities $Q(k' \mid k)$ for all the members, estimating the values from the data for all members combined. This simplification is not necessary in general but will be used for the example. With this assumption, the expected frequency of occurrence of a given diad reduces to

$$E(X) = [\tfrac{1}{2}Q(k_2 \mid k_1)Q(k_2' \mid k_1')]\,[\sum_i P_i(k_1) \sum_i P_i(k_1') - \sum_i P_i(k_1)P_i(k_1')]$$

and the expression for the variance is similarly simplified.

For our group, the conditional probabilities observed are shown in Table 3. If we now combine the diads as we did in Table 2, and compare the frequencies observed and the frequencies predicted using the conditional probabilities, we can observe a striking improvement in agreement. (Cf. Table 4 and compare with Table 2).

Using the chi-square test, we see that there is a probability of about 0.40 that the value of chi-square observed would be exceeded if the hypothesis were true. We can on this evidence neither accept nor reject the hypothesis that the observed frequencies of these diad types are accounted for by a chance model with predictions conditioned by feelings; but the improvement in fit is striking and suggests that a large part of the observed distribution of diad types is due to such contingency. This example illustrates the use of such baseline models in the study of the meaning of the observed frequency distribution for the diad types.

It is obvious that other hypotheses about the group could be tested by

constructing a similar model and examining the observed frequencies to determine how much of the variation is accounted for by such a model. The principle in all cases is the same; a model is constructed which assumes that the members of the group are automata acting at random, with probabilities governed by the particular hypotheses at hand. The expected frequencies obtained from this model are then used to investigate the group and to determine whether we have reason to believe that other psychological processes are at work beyond those assumed in the model. These hypotheses must be chosen with care, however, in order to yield a model which is mathematically tractable and which leads to a practical amount of computational labor.

III. *Simplifications for Large Groups*

In the models developed above, we have allowed the probabilities P_i and Q_i to be different for each member of the group. This leads to lengthy calculations for large groups. For groups larger than 10, however, we may introduce a simplification which greatly reduces this labor by using the mean value over all members of the group for the value of P_i ; thus each member is described by the same probabilities; thus summations are no longer necessary. In the case of the first model mentioned above, if the mean value of $P_i(k_1)$ is denoted by $P(k_1)$, and the mean of $Q_i(k_2)$ by $Q(k_2)$, we may then write the expected value $E(X)$ as

$$E'(X) = [n(n-1)/2][P(k_1)Q(k_2)P(k_1')Q(k_2')],$$

and the expression for the variance is similarly simplified.

Let us examine the error involved in this approximation. Let

$$A(X) = \sum_i P_i(k_1)Q_i(k_2) - nP(k_1)Q(k_2)$$

and

$$A'(X) = \sum_i P_i(k_1')Q_i(k_2') - nP(k_1')Q(k_2')$$

and

$$B(X) = \sum_i P_i(k_1)Q_i(k_2)P_i(k_1')Q_i(k_2') - nP(k_1)Q(k_2)P(k_1')Q(k_2').$$

Then it can easily be shown that if $E(X)$ is the expected value previously calculated using the individual probabilities, and $E'(X)$ is the expected value given above,

$$\begin{aligned} E(X) - E'(X) = \tfrac{1}{2}[&A(X)A'(X) + nA(X)P(k_1')Q(k_2') \\ &+ nA'(X)P(k_1)Q(k_2) - B(X)]; \end{aligned}$$

and so if we use $D(X) = [E(X) - E'(X)]/E'(X)$ as a measure of the error,

$$\begin{aligned} D(X) = &\frac{A(X)A'(X)}{n(n-1)P(k_1)Q(k_2)P(k_1')Q(k_2')} + \frac{A(X)}{(n-1)P(k_1)Q(k_2)} \\ &+ \frac{A'(X)}{(n-1)P(k_1')Q(k_2')} - \frac{B(X)}{n(n-1)P(k_1)Q(k_2)P(k_1')Q(k_2')}. \end{aligned}$$

Now it has been found from experience that $A(X)/[P(k_1)Q(k_2)]$ and $A'(X)/[P(k_1')Q(k_2')]$ are less than 2, and in almost all cases very near 1 for the groups encountered in practice. $B(X)/[(n-1)P(k_1)Q(k_2)P(k_1')Q(k_2')]$ is less than 1, and in almost all cases less than 1/2; so in practice this error $D(X)$ is less than $5/n$ for n greater than 5. In almost all cases this turns out to be a very liberal estimate of the error; for example, in the group of 10 used earlier, typical errors are

$$D[(00)(00)] = 0.00738 = .74\%$$

$$D[(00)(01)] = 2.28\%$$

$$D[(00)(02)] = 0.74\%$$

For the model with $Q_i(k_2)$ given by $Q(k_2 \mid k_1)$ for all members, the error in replacing the P_i by P is even less. In this case, A and $A'(X)$ are 0, and the error is then less than $1/n$.

This simplification is particularly useful because it introduces the least error for large groups, where it is most needed to simplify the calculation.

IV. *Summary*

Relational analysis defines the diadic relationship existing between pairs of members of a group in terms of their choices, their rejections, and their perceptions of being chosen and rejected. The number of possible diads is 45. In order to interpret the results of an experiment, we must have knowledge of the expected occurrence of the various diads on a chance basis, when only certain specified processes govern the chance distribution of diads.

This paper discusses the construction of models which give the expected value and variance of the diads, when certain assumptions are made as to the random factors operating in the group. In general, the assumptions are that only very simple psychological factors are operating in the group, and that the occurrence of the various diad forms is the result of chance operating within the restriction of these factors. The observed data are then examined to determine whether the chance model accounts for the distribution of diad types, or whether additional psychological processes must be postulated. Models such as these are essential for testing various hypotheses about interaction in the group since they provide a method for setting up and testing a null hypothesis by the usual statistical methods.

The models discussed in the paper were constructed on the assumption that choices and predictions were independent from member to member, and under the assumption that prediction was conditioned by choice in any pair as well as the assumption that choice and prediction were independent. The first of these assumptions was shown to account for a large part of the observed variation in diad frequency. Simplified assumptions which are valid for large groups were also discussed.

Models such as the second one discussed in this paper are typical of a large variety of models which could be constructed to test various hypotheses about the sources of the variation of frequency of the diad types.

REFERENCES

1. Feller, W. An introduction to probability theory and its application. Vol. 1. New York: Wiley, 1950.
2. Tagiuri, R. Relational analysis: An extension of sociometric method with emphasis upon social perception. *Sociometry*, 1952, **15**, 91–104.
3. Tagiuri, R., Blake, R. R., and Bruner, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.*, 1953, **48**, 585–592.

Manuscript received 7/7/53

Revised manuscript received 11/22/54

PSYCHOMETRIC SOCIETY

Statement of Receipts and Disbursements for Fiscal Year
Ended June 30, 1955

RECEIPTS (Dues)

Year	Members	Student Members
1957	1 ($7 each)	
1956	3 ($7 each)	
1955	432 ($7 each)	46 ($4 each)
1954	66 ($7 each)	19 ($4 each)
1953	5 ($7 each)	
	507	65

		$3809.00
Received with Dues for Corporation Publications		7.20
Mailing List Rental		25.00
Proceeds of 1954 Joint Dinner with APA Division 5		23.13
Miscellaneous		.15
Total Receipts		$3864.48

DISBURSEMENTS

Psychometric Corporation (90% of dues)		$3428.10
Miscellaneous Disbursements:		
Corporation for Publications	$ 7.20	
Mimeographing and Printing	67.77	
Postage	70.92	
Secretarial Services	165.02	
Addressing and Mailing	34.79	
Bank Charges	6.79	
		352.49
Total Disbursements		$3780.59

BALANCE

Reported Bank Balance, July 1, 1954	$1014.68
Actual Bank Balance, July 1, 1954	$1003.91
Receipts	3864.48
	$4868.39
Disbursements, 1954-1955	3780.59
Bank Balance, June 30, 1955	$1087.80

PSYCHOMETRIC CORPORATION

Statement of Receipts and Disbursements for Fiscal Year
Ended June 30, 1955

Reported Bank Balance, July 1, 1954		$6322.53
Actual Bank Balance, July 1, 1954		$6320.72

RECEIPTS

Subscriptions (less agency discounts)	$5439.50	
Psychometric Society (90% of dues)	3428.10	
Sale of Back Issues (less discounts)	1404.23	
Sale of Monographs 5-8 (less discounts)	546.00	
Royalties from U. Chi. Press (1953-1954)	1.38	
Reprints, Etchings, and Alterations	206.11	
For Monographs 2-4 (less discounts)	117.95	
Overpayments	12.35	
Total Receipts		$11155.92

DISBURSEMENTS

Printing Psychometrika, Volume 19, No. 2, through 20, No. 1	$5030.58	
Reprints, Etchings, and Alterations	197.86	
Stipend of Assistant Editor, Volume 18, No. 4, through 19, No. 2	246.00	
Secretarial Services	456.81	
Stationery and Postage	271.15	
Loan Repayments in Full	1100.00	
For Monographs 2-4	105.75	
Miscellaneous	75.60	
Total Disbursements		$7483.75

BANK BALANCE, June 30, 1955	$9992.89

ESTIMATED OBLIGATIONS

Printing Psychometrika, Volume 20, Nos. 2, 3, 4	$3900.00	
Stipend of Assistant Editor, Volume 19, No. 3, through 20, No. 4	544.00	
Secretarial Services	750.00	
Handling Charges on Back Issues and Monographs through 6/30/55	825.00	
Total Obligations		$6019.00

BANK BALANCE (less obligations), June 30, 1955	$3973.89

INDEX FOR VOLUME 20

AUTHOR

PSYCHOMETRIKA

Author and Subject Indexes

1946–1955

Prepared by B. J. Winer
Department of Psychology
Purdue University

AUTHOR INDEX

SUBJECT INDEX

In the selection of the articles to be printed in *Psychometrika*, an effort is made to obtain objectivity of choice. All manuscripts are received by one person, who first removes from each article the name of contributor and institution. The article is then sent to three or more persons who make independent judgments upon the suitability of the article submitted. This procedure seems to offer a possible mechanism for making judicious and fair selections.

Prospective authors are referred to the "Rules for Preparation of Manuscripts for Psychometrika," contained in the September, 1953 issue. Reprints of these "Rules" are available from the managing editor upon request. A manuscript which fails to comply with these requirements will be returned to the author for revision.

Authors will receive 100 reprints without covers, free of charge.

Manuscripts for publication in *Psychometrika* and material for review should be sent to

DOROTHY C. ADKINS, Managing Editor, *Psychometrika*
Department of Psychology, The University of North Carolina
Chapel Hill, North Carolina

The officers of the Psychometric Society for the year September 1955 to September 1956 are as follows: *President:* Clyde H. Coombs, Department of Psychology, University of Michigan, Ann Arbor, Michigan; *Secretary:* Lyle V. Jones, Department of Psychology, University of Chicago, Chicago 37, Illinois; *Treasurer:* David R. Saunders, Educational Testing Service, P. O. Box 592, Princeton, New Jersey.

The Council members, together with dates at which terms expire, are as follows: Oscar K. Buros, 1958; Alston S. Householder, 1958; Gertrude M. Cox, 1957; Harry H. Harman, 1957; Paul S. Dwyer, 1956; E. F. Lindquist, 1956.

The *Psychometric Monographs Committee* is composed of Frederick B. Davis, *Chairman;* Harold Gulliksen, Paul Horst, and Frederic Kuder. Manuscripts and correspondence for this series should be addressed to

FREDERICK B. DAVIS, Chairman, Psychometric Monographs Committee,
Hunter College, 695 Park Avenue, New York 21, N. Y.

PRINTED BY THE WILLIAM BYRD PRESS, INCORPORATED, RICHMOND, VA.